```
Db      617 --------GlnSerLysAsnLysGluThrGlyAspValLeuSerValThrProArgTyr 633
                    |||                              ||||||       :::

Qy     1711 ATGGTGAATCGAAACTGAACTGGCAGATCACCGAGAGTGGCATCATGGCTGGTGCC 1770
                    :::|||||:::  |||:||||:::|:::::::
Db      634 ThrLeuAsnSerMetLeuAspTrpGlnAlaThrAspAlaSerLeuGlnAlaThrVal 653

Qy     1771 CGTATCGCGGAAAACACCACGTTCACCCAGAATTATTCGTCACTGAGCGCTGTACAG 1830
                    :::    |||||||                              :::
Db      654 ThrTrpTyrGlyLysGln--------------------------------Lys 660

Qy     1831 AAGAAAGTGTATGATGAGAAAGGA-------------GAATACCTGAAA 1866
                    |||:||||||:::                       :::
Db      661 ProLysLysTyrAspTyrHisGlySerAlaAsnAspGlnLeuSer 680

Qy     1867 GCCTGGACGGTGGTGGTGCAGGTCGTGTCTGGAAGATGACGGATGCCCTGACGCTGAAT 1926
                    :::  |||:||||||||  :::::::::::||:||:||
Db      681 ProTyrAlaIleAlaGlyLeuGlyGlyThrTyrArgLeuSerLysAsnLeuSerLeuGly 700

Qy     1927 GCTGCGGTGAATAACCTGCTCAACAAGGATTACAGTGACGTGAGCCTGTACAGTGCCGGT 1986
                    ||:||||||| :::||||     |||:||:||
Db      701 AlaGlyValAlaAspAsnLeuPheAspLys----------------ArgLeuPheArgAlaGly 715

Qy     1987 AAGAGTACGCTGTATGCCCGTGATTACTTCCAGACGGGATCATCAACAACAGGATATGTG 2046
                    :::                              |||
Db      716 AsnAlaGlnGlyValValGly--------IleAspThrGlyAlaAlaThrTyrAsn 732

Qy     2047 ATACCTGAGCGAAATTACTGGATGTCGTGAACTATCAGTTC 2088
                    :::::::::                     |||
Db      733 GluProGlyArgThrPheTyrThrSerLeuThrAlaSerPhe 746


RESULT 5
BTUB_ECOLI
ID   BTUB_ECOLI   STANDARD;   PRT;   614 AA.
AC   P06129;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Vitamin B12 receptor precursor.
GN   BTUB OR BFE OR CER OR DCRC OR B3966.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85130824; PubMed=3882670;
RA   Heller K., Kadner R.J.;
RT   "Nucleotide sequence of the gene for the vitamin B12 receptor protein
RT   in the outer membrane of Escherichia coli.";
RL   J. Bacteriol. 161:904-908(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94089392; PubMed=8265357;
RA   Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA   Daniels D.L.;
RT   "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT   region from 89.2 to 92.8 minutes.";
RL   Nucleic Acids Res. 21:5408-5417(1993).
RN   [3]
RP   SEQUENCE OF 1-5 FROM N.A.
RX   MEDLINE=91154132; PubMed=1999392;
RA   Gustafsson C., Lindstroem P.H., Hagervall T.G., Esberg B.,
RA   Bjoerk G.R.;
RT   "The trmA promoter has regulatory features and sequence elements in
RT   common with the rRNA P1 promoter family of Escherichia coli.";
RL   J. Bacteriol. 173:1757-1764(1991).
RN   [4]
RP   SEQUENCE OF 456-614 FROM N.A.
RC   STRAIN=RDD020;
RX   MEDLINE=93106943; PubMed=8093236;
RA   Dougherty T.J., Thanassi J.A., Pucci M.J.;
```

```
RT   "The Escherichia coli mutant requiring D-glutamic acid is the result
RT   of mutations in two distinct genetic loci.";
RL   J. Bacteriol. 175:111-116(1993).
RN   [5]
RP   SEQUENCE OF 21-32.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE=97443975; PubMed=9228646;
RA   Link A.J., Robison K., Church G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [6]
RP   MUTAGENESIS OF TONB BOX.
RX   MEDLINE=90078094; PubMed=2687240;
RA   Gudmundsdottir A., Bell P.E., Lundrigan M.D., Bradbeer C.,
RA   Kadner R.J.;
RT   "Point mutations in a conserved region (TonB box) of Escherichia coli
RT   outer membrane protein BtuB affect vitamin B12 transport.";
RL   J. Bacteriol. 171:6526-6533(1989).
CC   -!- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR
CC       BACTERIOPHAGE BF23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M10112; AAA23524.1; -.
DR   EMBL; U00006; AAC43072.1; -.
DR   EMBL; AE000471; AAC76948.1; -.
DR   EMBL; M57568; -; NOT_ANNOTATED_CDS.
DR   EMBL; L14556; AAA23676.1; -.
DR   PIR; A65204; QRECBT.
DR   EcoGene; EG10126; btuB.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Outer membrane; Cobalt transport; Transport; TonB box; Signal;
KW   Phage recognition; Receptor; Complete proteome.
FT   SIGNAL        1     20                VITAMIN B12 RECEPTOR.
FT   CHAIN        21    614                TONB BOX.
FT   SITE         26     33                TONB C-TERMINAL BOX.
FT   SITE        597    614                L->P: INACTIVATES UPTAKE.
FT   MUTAGEN      28     28                V->G: INACTIVATES UPTAKE.
FT   MUTAGEN      30     30                A -> G (IN REF. 1).
FT   CONFLICT    162    162                A -> R (IN REF. 1).
FT   CONFLICT    377    377
SQ   SEQUENCE   614 AA;   68407 MW;   AB43CC46A991FF95 CRC64;
```

```
Alignment Scores:
Pred. No.:                                Length:              614
Score:                   2.33e-30         Matches:             187
Percent Similarity:      528.50           Conservative:        118
Best Local Similarity:   42.24%           Mismatches:          246
Query Match:             25.90%           Indels:              171
DB:                      14.16%           Gaps:                30
                         1

US-10-625-972-4 (1-2091) x BTUB_ECOLI (1-614)

Qy      19 GCTTCCGTAGTCATTCCCTGT------CTCGGATTTTCAGCCAGCAGCATAGCTGCTGCA 72
                 |||:|||:::             :::|||||||::   |||
Db       5 AlaSerLeuLeuthrAlaCysSerValThrAlaPheSerAlaTrpAlaGlnAspThrSer 24

Qy      73 GAGGATGATGTGATGATTGTTCGGCATCCGGCATGAGAAAAAGCTGACTAACGGACCGCC 132
                 :::|||||||||::::|||:::  :::||:||||||||:::
Db      25 ProAspThrLeuValValThrAlaAsnArgPheGluGlnProArgSerThrValLeuAla 44
```

Db 133 AGTGTTTCTGTGATTAGCCAGGAGGAATTGCAGTCCAGCCACGACCGATCTGGCGGAG 192
              :::::|||||:::::::||| ||: ||
Qy   45 ProThrThrValValThrArgGlnAspIleAspArgTrpGlnSerThrSerValAsnAsp 64

Qy 193 GCTCTGAGATCAGTAGAGGGTGTGATGTTGAAAGTGGTACGGGTAAAACCGGAGGGCTG 252
              |||||| ::: ||||||||| |||
Db   65 ValLeuArgArgLeuProGlyValAspIleThrGlyGlySerGlyGlnLeuSer 84

Db 253 GAAATCAGGATCCCAGGAGAATGCCAGCAGTTACACGCTGATACTGCATTGATGTGTTCGT 312
              :::|||||||::|::   ::: |||
Qy   85 SerIlePheIleArgGlyThrArgAsnAlaSerHisValLeuValLeuIleAspGlyValArg 104

Qy 313 -------CAGGGCAGAAGCAGTGACGTGACTCCCAACGGTTTTCTGCCATG 357
                   |||||||:::::|||
Db 105 LeuAsnLeuAlaGlyValSerGlySerArgIleAspLeuSer 117

Db 358 AATACCGGGTTCATGCCCCCTCTGGCGCCATTGAGCGTTATTGAGGTTATCAGGGGGCCG 417
              ::|:::|| :::::|||:::|||||||||
Qy 118 -------GlnPheProIleAlaLeuValGlnArgValGlnTyrIleArgGlyPro 133

Qy 418 ATGTCCACACTGTATGGCTCTGATGCGGTCGGTGCGTGGTGAATATCATTACCAGAAAG 477
              :::|||||||:::||| :::
Db 134 ArgSerAlaValTyrGlySerAspAlaArgIleGlyValGlyValValAsnIleIleThrArg 153

Db 478 AATGCAGACAAATGGCTCTCTTCCGTCAATGCAGGGCGTGAATCTGCAGGAAAGCAACAA 537
              |||:::   :::::::|||
Qy 154 ------AspGluProGlyThrGluIleSerGlyAlaGly- 163

Qy 538 TGGGGTAACAGCAGCCAGTTAATTTCTGGAGCAGTGGTCCCCTTGTGGATGATTCTGTC 597
              ||||||:::::::||||||
Db 164 TrpGlySerAsnSerTyrGlnAsnTyr- 172

Db 598 AGCCTGCAGGTACGCGGTAGCACAACAGCGTCAGGGT----TCATCGGTCACATCA 651
              |||||||||::||| ::: |||
Qy 173 ----------AspValSerThrGlnIleGlnLeuGlyAspIleTyrArgValThrLeu 188

Qy 652 CTGAGCGGATACAGCAGGCACGCGTATTCCTATCCCACCGAGTCCAGAGAATTATAAATCTT 711
              |||::: ||||||| |||
Db 189 LeuGlyAspTyrAlaHisThrHis- 196

Db 712 GGTGCACGTCTTGACTGGAAGGCGTCGGAGCAGGATGTGCTCTGGTTTGATATGGATACC 771
              ::: |||
Qy 197 -----------GlyTyrAspValValAlaAlaTyrGlyAsnThrGly 207

Qy 772 ACCCGGCAGGTTATGATAACCGGATGGCGGTATCTGACGGGGGGGATATGAC 831
              |||||::: |||||| ::: |||
Db 208 ThrGlnAlaGlnThrAspAsn--AspGlyPheLeuSerLysThrLeuTyrGly--- 224

Db 832 CGGACCCGTGCCCTATCGAGCGGAAACAAAATTTCAGCTGCTATGATCATCATTTCACCTTC 891
              ::: |||||||
Qy 225 -----------AlaLeuGluHisAsnPheThr-- 231

Qy 892 GGAACATGGAAATCGTATCTG---------AACTGGAACGAG 924
              :::||||||:::::::
Db 232 AspAlaTrpIleSerGlyPheValArgGlyTyrAspGlyTyrArgAsnArgThrAsnTyrAspAla 251

Db 925 ACAGAAAATAAAGGTCGTGAGCTTGTACCGCAGT--GTACTGAAGCGGCGACAAATGGGGG 981
              ::: :::|||:::|||:::
Qy 252 TyrTyrSerProGlySerProLeuLeuAspThrArgLysIleGlyTyrSerGlnSerTrpTrpAsp 271

Qy 982 CTTGCCGGTCAGCCGCGGGAGCTTAAGGAATGCAAACCTTATCCTGAATTCATTACTGCTT 1041
              ||||||||:::::::::::
Db 272 ---AlaGly-----------LeuArgTyrAsnGlyGluLeuIleLysSerGlnLeuIle 286

Db 1042 ACCCCCTTCGGGAGAATCT--------------CATCTGGTTACGGTGGGGGGC 1080
              :::::|||
Qy 287 ThrSerTyrSerHisSerLysAspTyrAsnTyrAspProHisTyr----------Gly 302

Qy 1081 GAGTTTCAGAGCTCGTCCATCGAAAGACGGAGTT--- ::: 1113
              ::: :::
Db 303 ArgTyrAspSerSerAlaThrLeuAsnpGluMetLysGlnLysIleTyrThrValGlnTrpAlaAsn 322

Qy 1114 ---GTCCTTGCCAGACAGGTGAAACTTTCCGGCAGAAAAGCTCG------- 1155

Db 323 AsnValIleValAlaGlyHisGlySerIleGlyAlaAlaGlyValAspGlyTrpGlnLysGlnThrThr 342
              ||||::
Qy 1156 -------TCGGTATTGCTGACGATGAGTGG- 1185
              ::: ||||||
Db 343 ThrProGlyThrGlyTyrValGluAspGlyThrTyrAspGlnArgAsnThrGlyIleTyrArgLeu 362

Qy 1186 ACGGATGCA------------CTTGCGCTGACTGCGGGCAGCGCCTATGAACATCAT 1230
              ||| ::: :::
Db 363 ThrGlyLeuGlnLnValGlyAspPheThrPheGluGlyIleAlaAlaArgSerAspAspAsn 382

Qy 1231 GAGCAATTCGGGGGACACTTCAGTCCGTCATATCGGTCTGGGATGTGGCAGATGCC 1290
              |||:::   :::::
Db 383 SerGlnPheGluGlyArgHisGlyThrTrpGlnThrSerAlaGlyTrpGluPheIleGluGlyLy 402

Qy 1291 TGGACGCGTGAAAGGCCGGTGACCAGGGGGATATAAGGCACCCAGAATGGGGCAGCTACAT 1350
              |||:::   :::||||||
Db 403 TyrArgPheIleAlaSerTyrGlyThrTyrLysSerAlaProAsnLeuGlyGlnLeuGlnLeuTyr 422

Qy 1351 AAAGGGATTAGTGGTGGTGTCCGGCAGGAGGAAAAACAAATCTACTTGGTAACCCCGACCTG 1410
              |||::: |||
Db 423 ------------------------GlyPheTyrGlyAsnProAsnLeu 430

Qy 1411 AAGCCCGGAAGAGAGCGTCAGTTGATGAGGCGGTGTATTACGATAACCCGGCCGGTCTG 1470
              |||:::|:::: :::
Db 431 AspProGluLysSerGlnTrpGlu---GlyAlaPheGluGlyLeuThrAlaGlyVal 449

Qy 1471 AATGCCAATGTCACAGGTTTATGACTGACTTCTCCAACAGAGTGTCTCTTATTCCATA 1530
              |||:::::: :::::: |||
Db 450 AsnTrpArgIleSerGlyTyrArgAsnAspValSerAsp---LeuIleAspTyr- 466

Qy 1531 AATGATAACACCAATAGCTATGTAAACAGCGGAAAGGCCCGGTTGCACGGTGTGGAATTT 1590
              |||:::|||||| :::
Db 467 AspAspHisThrLeuLysTyrTyrAsnGluGlyLysAlaArgIleLysGlyValGluAla 486

Qy 1591 GCCGGCCACATTGCGCGTGTGGTCAGAGGATGTCACGCTGTCACTGAATTATTACCTGACC 1650
              |||||||||||
Db 487 ThrAlaAsnPheAspThrGlyProLeuThrHisThrValSerTyrValSerTyrValAspAsnAla 506

Qy 1651 CGAAGTGAACAACGTGATGGTGATAACAAAGGTGCCGCCGTGAGTTATACCCCTGAACAC 1710
              |||::: ||||||
Db 507 ArgAsnAlaIleThrAsp-----------ThrProLeuLeuArgArgAlaLysGln 521

Qy 1711 ATGGTGAATGCGAAACTGAACTGCAGATCACCGACATCACCGAAGAGGTGGCATCATGGCTGGGTGCC 1770
              |||:::::::|||
Db 522 GlnValLysTyrGlnLeuAspTrpGlnLeuTyrAsp------PheAspTrp----GlyIle 538

Qy 1771 CGTTATCCGCGGGAAAACACCAGCTTTCACCCGGAGAATTATTCGTCACTGAGCCGCTGTACAG 1830
              |||||||||::: :::
Db 539 ThrTyrGlnTyrLeuGlyThrArgTyrArgTyrAspLysAspSerTyrSerSer- 553

Qy 1831 AAGAAAAGTGTATGATGAGAAAGGAGAATACTTGAAAGCCTGACGGTGGTGGATGCAGGT 1890
              ::: |||||||
Db 554 --------TyrProTyrGlnThrValLysMetGlyValGlyValSerLeuTrpAspLeuAla 570

Qy 1891 CTTGTCCGTGAAGATGACCGGATGCCCTGACGCGTGAATGCTGCGGTGAATAACCTGCTCAAC 1950
              :::::: :::::||
Db 571 ValAlaTyrProValThrSerHisLeuThrValArgGlyLysIleAlaAsnLeuProAsp 590

Qy 1951 AAGGATTACAGTGCACGTGAGCCTGTAC---AGTGCCGGTAAGACGTAAGAGTACGCGTGTATGCCGGT 2007
              ||||||::: :::|||
Db 591 LysAspTyrGluThrValTyrGlnThrValTyrTyrThrArgAlaGlyTyrThrLeuSerGly 610

Qy 2008 GATTAC 2013

Db 611 SerTyr 612

RESULT 6
BTUB_SALTY
ID   BTUB_SALTY   STANDARD;   PRT;   614 AA.
AC   P37409;
DT   01-OCT-1994 (Rel. 30, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)

```
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Vitamin B12 receptor precursor.
GN   BTUB OR STM4130.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=93079446; PubMed=1448622;
RA   Wei B.Y., Bradbeer C., Kadner R.J.;
RT   "Conserved structural and regulatory regions in the Salmonella
RT   typhimurium btuB gene for the outer membrane vitamin B12 transport
RT   protein.";
RL   Res. Microbiol. 143:459-466(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR
CC       BACTERIOPHAGE BF23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M89481; AAA27031.1; -.
DR   EMBL; AE008893; AAL22968.1; -.
DR   StyGene; SG10030; btuB.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Outer membrane; Cobalt transport; Transport; TonB box; Signal;
KW   Receptor; Complete proteome.
FT   SIGNAL        1     20
FT   CHAIN        21    614       VITAMIN B12 RECEPTOR.
FT   SITE         26     33       TONB BOX.
FT   SITE        597    614       TONB C-TERMINAL BOX.
FT   CONFLICT     81     81       G -> R (IN REF. 1).
FT   CONFLICT    134    134       R -> P (IN REF. 1).
FT   CONFLICT    183    183       N -> I (IN REF. 1).
FT   CONFLICT    280    280       R -> S (IN REF. 1).
SQ   SEQUENCE   614 AA;  68525 MW;  9F51F601A615FD62 CRC64;
```

Alignment Scores:

Pred. No.:	4.63e-29	Length:	614
Score:	510.50	Matches:	179
Percent Similarity:	43.06%	Conservative:	122
Best Local Similarity:	25.61%	Mismatches:	253
Query Match:	13.68%	Indels:	145
DB:	1	Gaps:	28

```
US-10-625-972-4 (1-2091) x BTUB_SALTY (1-614)

Qy    46 TTTTCAGCCAGCAGCATAGCTGCTGCAGAGGATGTGATTGATTGTCTGCGCATCCGGCTAT 105
            |||||||:  ||  :: |||  |  :::::::|::::::||:||| ||   :: ||
Db    16 PheSerAlaTrpAlaGlnAspThrSerProAspThrValLeuValThrAlaAsnArgPhe 35

Qy   106 GAGAAAAGCTGACTAACGCAGCCGCCAGTGTTCTGTGATTAGCCAGGAGGAATTGCAG 165
            ::::::           |||    :::::::::::::||
Db    36 GlnGlnProArgSerAlaValLeuAlaProValThrLeuValThrArgGlnIleGluGlu 55

Qy   166 TCCAGCCAGTACCACGATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAA 225
            ||  :: ::     || :: ::
Db    56 ArgTrpGlnSerThrSerValAsnAspValLeuArgArgLeuProGlyValAspIleAla 75

Qy   226 AGTGGTACGGGTAAAACCGGAGGGCTGGAAATCAGCATCCAGGAATGCCAGCCAGTAC 285
            |||||||||||
Db    76 GlnSerGlyGlyAlaGlyGlnAsnSerSerIlePheIleArgGlyThrAsnSerSerHis 95

Qy   286 ACGGTCGATACTGATTGATGGTGTTCGT------------CAGGGCGGAAGCAGTGAC 330
            ||||||:||||||||||
Db    96 ValLeuValLeuIleAspGlyValArgLeuAsnLeuAlaGlyValSerGlySerAlaAsp 115

Qy   331 GTGACTCCCAACGGTTTTCTGCCATGAATACCGGGTTCATGCCCCTCTGGCCGCCATT 390
            |||||||||
Db   116 GlnPheProValSerLeuVal 124

Qy   391 GAGCGGTATTGAGGTTATCAGGGGGCCCATGTCCACACTGTAGGCCTCTGACGTCGATGGC 450
            :::|||||||||||||||||||||::|||||
Db   125 GlnArgIleGluTyrIleArgGlyProArgSerAlaIlePheTyrGlySerAspAlaIleGly 144

Qy   451 GGTGTGGTGAATATCATTACCAGAAAGAATGCAGAAAGAAATGGCTCTCTTCCGTCAATGCA 510
            ||||||||||||||||||||:::
Db   145 GlyValValAsnIleIleIleThrThrArgAsp-- :::::: AsnPro 156

Qy   511 GGGCTGAATCTGCAGAAAGCAACAAATGGGTAACAGCAGCCAGTTAATTTCTGGAGC 570
            :::|||||||||::::||
Db   157 GlyThrGluLeuThrAlaGly------TrpGlySerAsnSerTyrGlnAsnTyr--- 172

Qy   571 AGTGGTCCCCTTGTGGATGATTCTGTCAGCCTGCAGGTACGGCGGTAGCACACAACAGCGT 630
            |||||||||::::||
Db   173 ------------------------------AspIleSerThrGlnGlnGln 179

Qy   631 CAGGGTTCATCGGTCACATCACTGAGCGATACAGCAGCGCACGGTATTCCTATCCCACG 690
            :::|||
Db   180 ------------------GluAsnThrArgAlaThrLeuIle--- 189

Qy   691 GAGTCACAGAATTATAATCTTGGTGCACGTCTTGACTGGAAGGCGTCGGAGCAGGATGTG 750
            |||
Db   190 GlyAspTyrGluTyrThrLysGly--- 197

Qy   751 CTCTGGTTGATATG----------------GATACCACCCCGGCAGCGTTATGATAACCGGGAT 798
            |||||:::
Db   198 ------PheAspValAlaAlaLysGlyGlyValGlyThrGlyMetGlnAlaGlnProAspArgAsp 215

Qy   799 GGGCAACTGGGG--AGTCTGACGGGGGGGATAGACCGGACCCTG---CGCTATGAG 849
            |||   ||    : ::|||
Db   216 GlyPheLeuSerLysThrArgGlyPheAlaLeuGluHisThrPheSerAspArgTrpSer 235

Qy   850 CGAAACAAAATTTCAGCGGCTATGATCATCACTTTCACCTTCGGA 894
            |||||||||::: :::
Db   236 GlyPheValArgGlyTyrGlyTyrAspAspAsnArgThrAspTyrAspAlaTyrTyrSerPro 255

Qy   895 ----------------------ACATGGAAATCGTATCTGAAC 915
            ::::|| :::
Db   256 GlySerProLeuIleAspThrArgLysLeuTyrSerGlnSerTyrArgTrpAspAlaGlyLeuHis 275

Qy   916 TGGAACGAGACAGACAGAAAATAAAGTCGTGAGCTTGTACGGTACTG-- 972
            :::|||    ||| :::|||
Db   276 PheAsn---GlyGluArgIleGluGlnSerGlnLeuValSerSerTyrSerHisSerLysAsp 294

Qy   973 ------AAGCGGGACTTGCGGTACAGCCGCGGGAGCTTAAG 1008
            :::::: :::
Db   295 TyrAsnTyrAspProHisTyrGlyArgGlyTyrAspThrSerAlaThrLeuAspGluMetLys 314

Qy  1009 GAATCGAACCTTATCCTGAATTCATTACTGCTTACCCCTCTGGAGAATCTCATCTGTT 1068
            :::  ||| :::::::::||| ||
Db   315 GlnTyrAsnValGlnTrpThrAsnSerValValValGlyHisGlyAsn------Val 331
```

```
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
CC   -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE000242; AAC74533.1; -.
DR   PIR; F64897; F64897.
DR   EcoGene; EG13774; yncD.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW   Hypothetical protein; Outer membrane; Receptor; Signal; TonB box;
KW   Complete proteome.
FT   SIGNAL        1     23        POTENTIAL.
FT   CHAIN        24    700        PROBABLE TONB-DEPENDENT RECEPTOR YNCD.
FT   SITE        680    680        TONB C-TERMINAL BOX.
SQ   SEQUENCE   700 AA;  77260 MW;  7B3B96C6ABA48FE4 CRC64;
```

```
Alignment Scores:
Pred. No.:                        3.65e-19        Length:           700
Score:                            373.50          Matches:          186
Percent Similarity:               35.04%          Conservative:     109
Best Local Similarity:            22.09%          Mismatches:       250
Query Match:                      10.01%          Indels:           297
DB:                               1               Gaps:              37

US-10-625-972-4 (1-2091) x YNCD_ECOLI (1-700)
```

```
Qy    1 ATGTCGAATAACAACCACTCTGGCTTCCGTAGTCATTCCTGT--CTCCGGATTTCAGCCAGC   57
                 ||||||:|::|::|:|:|:|:::::::|:|:||     :::|
Db    1 MetLysIlePheSerValArgGlnThrValLeuProAlaLeuLeuValLeuSerProVal1   20

Qy   58 AGATAGTGCTGCAGAGGATGTGATGATTGTCTCCGGCATCCGATGAGAAAAAGCTG  117
                ||||||||||||||||:||||||||:||||||:|||::|::|:|::|
Db   21 ValPheAlaAlaAspGluGlnThrMetIleValSerAlaAlaProGlnValValSerGlu   40

Qy  118 ACTACAGCCAGCCGCCAGTGTTTCTGTGATTAGCCAGGAGAATTGCAG---TCCAGCCAG  174
                  |||::||:||:|:|:::|:|||  ||      |||  :::|
Db   41 LeuAspThrProAlaAlaValSerValAlaAspGlyGluMetArgLeuAlaThrPro   60

Qy  175 TACCACGATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTACG  234
                ||::|||:|::|:||::|::|::|||
Db   61 ArgIleAsnLeuSerGluSerLeuThrGlyValProGlyLeuGlnValGlnAsnArgGln   80

Qy  235 GGTAAAACCCGGAGGGCTGGAAATCAGCATCCGAGAATGCCAGCCAGTTACACGCTG---  291
                  ::: |:::|:|::|:|:|:::::|:|::|:|:::
Db   81 AsnTyrAlaGlnAsnAspLeuGlnLeuSerIleArgGlyPheGlySerThrTyrTyrGly  100

Qy  292 ------------ATACTGATTGATGGTGTT-------------CGTCAGGGGC  318
                           |||||||:|::|:                :|:|||:|
Db  101 IleArgGlyIleArgLeuTyrValAspGlyIleProAlaThrMetProAspGlyGlnGlyGly  120
```

```
Qy 1069 ACGGTGGGGGGCGAGTTCAGAGCTCGTCATGAAAGACGGAGTTGTCCTGCCAGCACA 1128
                 :::::||||||              :::
Db  332 GlyAlaAlaGlyLeuValAlaAspTrpGlnLysGlnThrThrProGly-------Thr  346

Qy 1129 GGT--------GAAACTTTCCGGCAGAAAAGCTGGTCGTATTGCTGAGGATGAGTGG 1179
                         ||||||||:|:   :::||:||
Db  347 GlyTyrValProGluGlyTyrAspGlnArgAsnThrGlyValTyrLeuThrGlyLeuGln  366

Qy 1180 CATCTCACGGATGCACTGCCTGCCAGGCGCTATGAACATCATGAGCAATTC 1239
                |||  |||    ::|||   ::|:||  ::: :||
Db  367 GlnLeuGlyAsp---PheThrLeuGluAlaAlaAlaArgSerAspAsnSerGlnPhe  385

Qy 1240 GGGGGACACTTCAGTCCGCGTGCATATCTGGTCTGGGATGTGGCAGATGCCTGACGCTG 1299
                |||||        :::|||  :::        :::
Db  386 GlyArgHisGlyThrTrpGlnThrSerAlaGlyThrTrpGluHisGluGluGlyTyrArgPhe  405

Qy 1300 AAAGGCGGTGTGACCACGGATATAAGGCACCCAGAATGGGACGCAGCTACATAAAGGGATT 1359
                |||||||||||||||:|:|:||:
Db  406 IleAlaSerTyrGlyThrSerTyrLysAlaProAsnLeuGlyGlnLeuTyr-------  422

Qy 1360 AGTGGTGTGTCCGGGCAGGGAAAAACAAATCTACTTGGTAACCCGACCTGAAGCCGGAA 1419
                ||||||||:::|||   ||||::::|
Db  423 ----------GlyTyrTyrGlyTyrAsnLeuAsnProGluGlu  433

Qy 1420 GAGAGGCCTCAGTTATGAGGTGTATTACGATAACCCCGGTCGAATGCCAAT 1479
                :::||  :::|||:::|:::
Db  434 LysSerLysGlnLeuThr---GlyAlaPheGluGlyLeuThrArgLeuGlyValSerTrpArg  452

Qy 1480 GTCACAGGTTTATGACTGACTTCTCCAACAAGATTGTCTTTATTCCATAAATGATAAC 1539
                :::::||||:|:   ::::||:||
Db  453 IleSerGlyTyrArgArgAsnAspIle----AsnAspMetIleAspTyr------AspAspHis  469

Qy 1540 ACCAATAGCTATGTAAACAGCGGAGAGGCCCGGTTGCACGGTGTGGAATTGCCGGCACA 1599
                |||||:|:|||:|:::|||:||||
Db  470 LeuGlnLeuLysTyrTyrAsnGluGlyLysAlaArglleLysGlyLleuGlnAlaThrAlaAsn  489

Qy 1600 TTGCCGCTGGTGGTCAGGAGGATGTCACGGTGTCACCGTGAATTACAACCTGGACCCGAAGTGAA 1659
                ||||||||:::||:||::|:|
Db  490 PheAspThrGlyProLeuThrHisThrValSerTyrArgAspTyrValAspAlaArgAsnAla  509

Qy 1660 CAACGTGATGCTGATAACAAGTGCGCCCGCTGAGTTATACCCCTGAACACATGGTGAAT 1719
                ||||
Db  510 IleThrAsp------------ThrProLeuProArgArgSerLysGlnMetAlaLys  524

Qy 1720 GCGGAAACTGAACTGGCAGATCACCGAAGAGGTGCATCATGGGTGGGGTGCCCGTTATCGC 1779
                :::||:::|:|::   :::   :::|||:
Db  525 TyrGlnLeuAspTrpAspValTyrAsp------PheAspTrp---GlyMetThrTyrGln  541

Qy 1780 GGGAAAACACCACGTTTCACCGAGATTATTCGTCACTGACGCGCTGTACAGAAGAAAGTG 1839
                :::|||||||||:::|||||:
Db  542 TyrLeuGlySerArgTyrArgAspSerAspTyrSerAla----------  553

Qy 1840 TATGATGAGAAAGGAGAATACCTGAAAGCCTGGACGGTGGTGGATGCGAGGTCGTGTCCGTGG 1899
                :::||:::|::||:|:||||:
Db  554 TyrProTyrArgThrValLysMetGlyValGlyValSerLeuTrpAspLeuTrpArgValAlaTyr  573

Qy 1900 AAGATGACGGATGCCCTGACGCGTGAATGCTGCGGTGAATAACCTGCTCAACAAGGATTAC 1959
                |||||||||||:::||:::
Db  574 ProValThrSerHisLeuThrValArgGlyLysIleLeuAlaAsnLeuPheAspLysAspTyr  593

Qy 1960 AGTGACGTGAGCCCTGTAC---AGTCCCGGTAAGAGTACCGCTGTATGCCGTGATTAC 2013
                :::||:::   ::::::||:|||:::|
Db  594 GluThrValTyrGlyTyrGlnThrAlaGlyTyrArgGlyGluTyrThrLeuSerGlySerTyr  612
```

```
RESULT 7
YNCD_ECOLI
ID   YNCD_ECOLI     STANDARD;     PRT;     700 AA.
AC   P76115;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable tonB-dependent receptor yncD precursor.
GN   YNCD OR B1451.
```

```
Qy  1221 GAGGATGAGTGGCATCTCACGGATGCACTTGGCGTCGACTGCGGGCAGCCGCTAT------ 1221
         ::: ::||||| ||: :::|| |||:|||
Db   402 GlnThrGlnTrpGlnLeuSerGluLysLeuAspAlaGlyLyrValArgTyrSerSer      421

Qy  1222 ------------------GAACATCATGAGCAATTCCGGGGA                     1245
         ::||||:::
Db   422 ValTrpPheAspSerAsnAspHisTyrValThrProGlyAsnGlyAspAspSerGlyAsp  441

Qy  1246 ---------CACTTCAGTCCGGTGCATATCTGGTCTCGGGATGTGGCAGATGCCTGG     1293
         ||||||||||:
Db   442 AlaSerTyrHisLysTrpLeuProAlaGlySerLeuLysTyrAlaMetThrAspAlaTrp  461

Qy  1294 ACGCTGAAAGGCGTGTGACCAATCTGCAGGAAGCAACAAATGGGGCAGCTACATAAA     1353
         ::: :||||
Db   462 AsnIleTyrLeuAlaAlaGlyTyrArgGlyPheGluThrProThrIleAsnGluLeuSerTyr  481

Qy  1354 CGGATTAGTGGTGTGTCCGGCAGGGAAAAACAAATCACTTGGTAACCCGACCTGAAG     1413
         |||||||
Db   482 ArgAlaAspGlyGlnSerGlyMet------AsnLeuGlyLeuLys                 494

Qy  1414 CCGGAAGAGAGCCTCAGTTATGAGGGTGTATTACGATAACCCCGGGTCTGAAT        1473
         |||||
Db   495 ProSerThrAsnAspThrIleGlyIleGlySerLysThrArgIleGlyAspGlyLeu---  513

Qy  1474 GCCAATGTCACAGGTTTTATGACTGACTTCTCCAACAAGATTGTCTCTTATTCCATAAAT  1533
         ::::|||||
Db   514 LeuSerLeuAlaLeuPheGlnThrArgAspThrAspAspLeuLeuValValAspSerSerSer  533

Qy  1534 GATAACACCAATAGCTATGTAAACAGCGGAGAAGGCCCCGGTTGCACGGTGTGGAATTGCC  1593
         ||||||
Db   534 GlyGlyArgThrThrTyrLysAsnAlaGlyLysThrArgGlnGlnAlaGlyLeuAla     553

Qy  1594 GGCACAATTGCCCGTGCGTGGTCAGAG------GATGTCACGCCTGTCACTGAATTAC    1641
         :::::|| :::::::
Db   554 ------------------TrpAspGlnArgPheAlaGlyArgValAsnAlaSerTrp     568

Qy  1642 ACCTGGACCCGAAGTGAACAACGTGAT---                                1668
         |||||
Db   569 ThrTrpLeuAspAlaThrTyrArgSerAsnValCysAsnGluGlnAspCysAsnGlyAsn  588

Qy  1669 ------------------GGTGATAACAAAGT--GCGCCGTGAGTTATACCCCTGAACAC  1710
         :::
Db   589 ArgMetProGlyIleAlaIleAlaArgAsnMetGlyPheAlaSerIleGlyTyrValProGluAsp  608

Qy  1711 ATGGTGAATGCGAAACTGAACTGGCACAGATCACCGAAGAGGTGCATCATGGCGTGGGTGCC  1770
         |||
Db   609 GlyTrpTyrTyrAlaGlyThrGlu------------------Ala                 616

Qy  1771 CGTTATGCGGG---------AAAAACACCACGTTTCACC                       1800
         |||
Db   617 ArgTyrMetGlyAspIleMetAlaAspGluAsnThrAlaLysAlaProSerTyrThr     636

Qy  1801 ------------------CAGAATTATTCGTCACTGAGCGGCCTGTACAGAAG         1833
         ::|||
Db   637 LeuValGlyLeuPheThrGlyTyrLysTyrHisAsnLeuThr--------           652

Qy  1834 AAAGTGTATGATGAGAAAGGAGAGAATACCTGAAAGCCTGAACGCCTGAATGCCATGCCTG  1893
         |||
Db   652 ------------------GlyTyrGlyLeuValAsnGluSerAsn---------GlyArgTyr  680

Qy  1894 TCCGTGGAAGATGACGGATGCCCTGACGTGAATGCTGCGGTGAATAACCTGCTCAACAAG   1953
         |||::||||||
Db   665 ------------------ValAspLeuProGlyAlaArgValAspAsnLeuPheAspLys   665

Qy  1954 GATTACAGTGACGTGAGCCTGTACAGTGCCGGTAAGAGTACCGTGATGCCGGTGATTAC   2013
         ::: :::
Db   666 GluTyr--------ValGlySerValIleValAsnGluSerAsn                 680

Qy  2014 TTCCAGACGGGATCATCAACAACAGGATATGTGATAACCTGAGCGGGAAATTACTGGATGTCG  2073
         :::::
Db   681 TyrGluProSer------ProGlyArgAsnTyrGlyValGly                   692

Qy  2074 CTGAAC 2079
```

```
Qy   319 GGAAGCAGTGACCTGACTCCCCAACGGTTTTCTGCCATGAAATACCGGGTTCATGCCCCCT  378
         :::|||::||:::::
Db   121 GlnThrSerAsnIle------------------------Asp                   126

Qy   379 CTGGCCGCCATTGAGCGTATTGAGGCGGCCGATGTCCCACTGTATGCTTCT           438
         |||:::||||||:||||:||
Db   127 LeuSerSerValGlnAsnAlaGluValLeuArgGlyTyrProPheSerAlaLeuTyrGly---  145

Qy   439 GATGCGATGGCGGTGTCGTGGTGATATCATTACCAGAAGAATGCAGACAAATGGCTCTCT  498
         ::||| |||||||::||:::
Db   146 AsnAlaSerGlyGlyValMetAsnValThrThrGlnInThrGlyGlnGlnPro-------- 162

Qy   499 TCCGTCAATGCAGGGCTGAATCTGCAGGAAGCAACAAATGGGTAACAGCAGCCAGTT     558
         ::: ||||||:::||:::
Db   163 ----------ProThrIleGluAlaSerSerTyrTyrGlySer--Phe           174

Qy   559 AATTTCTGGAGCAGTGGTCCCCTTGTGATGATTCTGTCAGCCAGGTACGCGGTAGC     618
         |||:::|||
Db   175 GlySerTrpArgTyrGly------LeuLysAlaThrGlyAla                 186

Qy   619 ACACAACAG------CGTCAGGGTCATCGTGCCACATCACTGAGCGGATACAGCAGGCACG 672
         ||| :::
Db   187 ThrGlyAspGlyProGlyAspValAspTyrThrValSerThrThrArgPheThr     206

Qy   673 CGTATTCCTTATCCCACGGAGTCA----------------CAGAATTATAAATCTTGGT 714
         |||
Db   207 ThrHisGlyTyrArgAspHisSerGlyAlaGlnLysAsnLeuAlaAsnLeuAlaLysLeuGly  226

Qy   715 GCACGTCTTGAC--------AACAAAATTCAGCTGGC---               870
         :::|||||
Db   227 ValArgIleAspGluAlaSerLysLeuSerLeuIlePheAsnSerValAspIleLysLysAla  246

Qy   727 ------------------TGGAAGGCGTCGGAGCAGCAGGATGTGCTCTGG          756
         |||||||||||:::
Db   247 AspAspProGlyGlyLeuThrLysAlaGluThrLysLysAlaAsnProGlnGlnAlaProArg  266

Qy   757 TTTGATATGGGATACCACCCGGCAG-------------CGTTATGAT             789
         |||||||:::::
Db   267 AlaGluGlnTyrAspThrArgLysThrIleLysGlnInThrAlaGlyLeuArgTyrGlu    286

Qy   790 ------------AACCGGCGGATGGCAACTGGGAGTCTGACGGGGGGATGACCGGACC    837
         ::: |||||||
Db   287 ArgSerLeuSerSerArgAspAspMetSerValMetMetTyrArgAlaGlyGluArgGluThr  306

Qy   838 CTGGCCTATGAGCGA-----                                          870
         :::||:||
Db   307 ThrGlnInTyrGlnSerIleProMetAlaProGlnLeuAsnProSerHisAlaGlyGlyValVal  326

Qy   871 -------------------TATGATCAT---                               879
         |||
Db   327 IleThrLeuGlnArgHisTyrGlnGlyIleLeuAspSerArgTrpThrHisArgGlyGluLeu  346

Qy   880 ------ACTTTCACCTTCGAACATGGAAATGCGTATCGAACTGGAACTGGAACAGCAGAAAT  933
         |||:::||||||:::
Db   347 GlyValProValThrPheThrThrGlyLeuAsnTyrGluAsnMetSerGlu--AsnArg   365

Qy   934 AAAGGTCGTCGAGCTTGTACCCAGTCTACTGTGACGTCGAAGCCGGACAAATGGGGGCTTGCCCGGTCAG  993
         :::::|||||
Db   366 LysGlyGlyTyrAsnAsnPheArgLeuAsnSerGlyMetProGluTyrGlyLnLysGlyGlu  385

Qy   994 CCGCGGGAGCTTAAGGAATCGAACCTTATCCTGAATTCATTACTGCTTACCCCCTCTGGGA  1053
         |||
Db   386 LeuArgArg--AspGluArgGlyAsnLeuMet--                           394

Qy  1054 GAATTCCATCTGGTTACGGTGGGGGGGCGAGTTTCAGAGCTCGTCCATGAAAGACGGAGTT  1113
         :::|||
Db   394 ------------------GlyValProValThrPheThrThrGlyLeuAsnMetSerGlu--AsnArg  394

Qy  1114 GTCCTTGCCAGCAGCAGGTGAAACTTTCCGGCAGAAAAAGCTGGTCGGTA-----TTTGCT  1167
         ||||||:::::
Db   394 ------------------TrpAsnIleAspProTyrLeu                       401

Qy  1168 GTCTTGCCAGCAGCAGGTGAAACTTTCCGGCAGAAAAAGCTGGTCGGTA            395
```

```
Db  693 MetAsn 694
            :::|||

Qy  109 AAAAGCTGACTACGACGCCCAGTGTTTCTGTGATTAGCAGGAGGAATTGCAGTCC 168
            |||:||||::|:||   |||:::|||
Db   55 ArgSerSerPheGluAlaProMetMetValThrValGluAlaAspThrProThrSer 74

Qy  169 AGCCAGTACCACGATCTGGCGAGGCTCTGAGATCAGTAGAGAGGGTGTTGGATGTTGAAAGT 228
            :::|||   ||||||| |||:::|    :::
Db   75 GluThrAlaThrSerAlaThrAspMetLeuArgAsnIleProGlyLeuThrVal---Thr 93

Qy  229 GGTACGGGTAAAACCGGAGAGGGTGGAAATCAGCATCCGAGGAATGCCAGCCAGTTACACG 288
            ||:::||||:::   |||   :::::::|||||
Db   94 GlySerGlyArgValAsnGlyGlnAspValThrLeuArgGlyTyrGlyLysGlnGlyVal 113

Qy  289 CTGATACTGATTGATGGTGTTCGTCAGGGCGGAAGCAGTGACGTGACTCCCAACGGTTTT 348
            |||   |||::|   :::
Db  114 LeuThrLeuValAspGlyIleLeuArgGlnGly--------ThrAspThrGlyHis 128

Qy  349 TCTGCCATGAATACCGGGTTCATGCCCCCTCGCCGCCATTGAGCGTATTGAGGTTATC 408
            :::||:||   |||:::  |||::::|||:::
Db  129 ------LeuAsnSerThrPheLeuAspPro---AlaLeuValLysArgValGluIleVal 145

Qy  409 AGGGGGCCGATGTCCACACTGTATGGCTCTGATGCGATGGGCGGTGGTGGTGAATATCATT 468
            ||||||   |||:::||:::|||::||||||||:::::::
Db  146 ArgGlyProSerAlaLeuLeuTyrGlySerGlyAlaLeuGlyValIleSerTyrTyrGlu 165

Qy  469 ACCAGAAAGAAATGCAGACAAAATGGCTCTTCCGTCAATGCAGCGGCTGAAATCGCAGGAA 528
            |||||||:::||:::||||
Db  166 ThrValAspAlaAlaAspLeuLeuLeuProGlyGlnAsnSerGlyTyrArgValTyrSer 185

Qy  529 AGCACAAATGGGGTAACAGCAGCCAGTTAATTTCGGAGCAGTGT---------------- 576
            |||:::    |||:::   ::||||
Db  186 AlaAlaAlaThrGlyAspHisSer---PheGlyLeuGlyAlaSerAlaPheGlyArgThr 204

Qy  577 CCCCTTGTGGATGATTCGTCAGCCTGCAGGTACGC------GGTAGCACACAACAGCGT 630
            ||||||:::   :::   :::
Db  205 AspAspValAspGlyIleLeuSerPheGlyThrArgAspIleGlyAsnIleArgGlnSer 224

Qy  631 CAGGGTTCATCGTCACATCACTGAGCGATACGACGCAGGCAGCGCGTATTCCTATCCCACG 690
            |||   :::   |||:::   |||
Db  225 AspGlyPheAsnAla--------------------ProAsn 231

Qy  691 GAGTCACGAGAATTATATCTTGGTGCACGTCTTGACTGGGAGAGGCGTCGGAGCCAGGATGTG 750
            :::   |||:::
Db  232 AspGluThrIleSerAsnValLeuAlaLysGlyThrTrpArgIleAspGlnIleGlnSer 251

Qy  751 CTCTGGTTTGATATGATGATCACCACCCGCAGCGGTTGATGATAACCGGGATGGG------ 801
            |||   :::
Db  252 Leu------------------SerAlaAsnLeuArgTyrTyrAsnAsnSerAlaLeuGluLeuPro 266

Qy  802 ------------------------------CAACTGGGAGTCTGACGGGG 822
                                          |||
Db  267 LysAsnProGlnThrSerAlaAlaSerSerThrAsnLeuMetThrAspArgSerThrIle 286

Qy  823 GGATATGACCGGACCCTGCGCTATGAGCGAAACAAATTCAGCTGGCTATGATCATACT 882
            :::||||||   |||   :::
Db  287 GlnArgAspAlaGlnLeuGlyLysTyrAsnIleLysProLeu-------AspGlnGlu 302

Qy  883 TTCACCTTCGGAACATGGAAATCGTATCTGAACTGAACTGAGAGACAGAGAGAAAATAAAGGTCGT 942
            :::|||                  :::::
Db  303 TrpLeuAsnAlaThrAlaGlnValTyr------TyrSerGluValGluIleAsnAlaArg 320

Qy  943 GAGCTTGTGACGCAGTGTACTGAGAGCGCGACAAATGGGGGCTTGCCGGTCAGCCGCGGGAG 1002
            |||   |||
Db  321 Pro--------------------GlnGlyThrProGluGlu 327

Qy 1003 ----------------------CTTAAGGAATGCAGAACCTTATCCTGAATTCATTACTTGCTTACCCCCT 1047
                                  |||              :::
Db  328 GlyArgLysGlnThrThrLysGlyGlyLysLeuGluAsnArgThrArgLeuPheThrAsp 347

Qy 1048 CTGGGAGAATCTCATCTGGTTACGGTGGGGGCGGAGTTCAGAGCTCGTCCATGAAAGAC 1107
            :::|||                              :::
Db  348 SerPheAlaSerHisLeuLeuThrTyrGlyThrGlu------AlaTyrLysGln 363
```

Alignment Scores:

Pred. No.:	7.82e-18
Score:	355.00
Percent Similarity:	37.96%
Best Local Similarity:	23.54%
Query Match:	9.51%
DB:	

Length:	687
Matches:	355
Conservative:	106
Mismatches:	284
Indels:	172
Gaps:	34

Receptor.

US-10-625-972-4 (1-2091) x HEMR_YEREN (1-687)

```
Qy   22 TCCGTAGTCATTCCCTGT-------CTCGGATTTCAGCCAGCAGCATAGCTGCTGCA 72
            ||||   |||  |||         |||:::      :::::::
Db   15 SerLeuAlaIleAlaCysThrLeuSerLeuAlaValGlnAlaAlaAspThrSerSerThr 34

Qy   73 GAG----------GATGTGATGATTGCTCGGCATCCGGCTATGAG 108
            :::        |||:||||:||::||||||||||||
Db   35 GlnThrAsnSerLysLysArgIleAlaAspThrMetValThrAlaAspThrGlyAsnGlu 54
```

```
Qy   1108  GGAGTTGTCCTTGCCAGCACAGGTGAAACTTTCCGGCAG--------AAAAGTCTGGTCG  1158
                 ::   ||||||||   :::
Db    364  GluGlnThrProSerGlyAlaThrGluSerPheProGlnAlaAspIleArgPheLysSer  383

Qy   1159  GTATTTGCTGAGGATGAGTGGCATCTCACGGAT--GCACTTGCGCTGACTGCGGGCAGC  1215
                 :::  :||||||||||       :::::::  ||||||||
Db    384  GlyTrpLeuGlnAspGluIleThrLeuArgAspLeuProValSerIleLeuAlaGlyThr  403

Qy   1216  CGCTATGAACATCATGAGCAACTTCGGGGGACACTTCAGTCCGCGTGCATATCTGGTCTGG  1275
                 |||||||::::::::                |||
Db    404  ArgTyrArgAspAsnTyrArgGly----------SerSerGluGlyTyrAlaAspVal  418

Qy   1276  GATGTGGCAGATGCCTGACGCTGAAAGGCGGTGTGACCACG------------------  1317
               ||  ||||||  ||||::  :::||||    ||||
Db    419  Asp--AlaAspLysTrpSerSerArgGlyAlaValSerValThrProThrAspTrpLeu  437

Qy   1318  ----------------GGATATAAGGCACCCAGAATGGGGCAGCTACAT--  1350
                            :::::||||||  ||||:|::|||
Db    438  MetLeuPheGlySerTyrArgAlaGlnAlaPheArgAlaProThrMetGlyGluMetTyrAsn  457

Qy   1351  ------AAAGGGATTAGT---GGTGTGTCCGGCAGGGAAAACAAATCTA---CTTGGT  1398
                  :::||||||||||:  :::::::||||
Db    458  AspSerLysHisPheSerMetAsnIleMetGlyAsnThrLeuThrAsnTyrTrpValPro  477

Qy   1399  AACCCCGACCTGAAGCCGGAAGAGAGCCAGTTATGGAGGCTGGGGTGATTACGATAAC  1458
               ||||||::::||||||||||    :::  ||
Db    478  AsnProAsnLeuLysProGluThrAsnGluThrGlnGluTyrGlyPhe----------  493

Qy   1459  CCCCCGGTCTGAATGCCAATGTCACAGGTTTATGACTGACTTCTCCAACAAGATTGTC  1518
                 ||||||   |||||       ||||||||:||||
Db    494  ----------GlyLeuArgPheAsnAspLeuMetMetAlaGluAspLeuGlnPheLysAla  511

Qy   1519  TCTTATTCCATAAATGATAACACCAATAGCTATGTAAACAGCGGAAACGCCGGTTGCAC  1578
               ||||||||||||||||||||||:||||
Db    512  SerTyr---PheAspThrAsnAlaLysAspTyrIleSerThrGlyValThrMetAspPhe  530

Qy   1579  GGTGTGGAATTTGCCGGC----------------------  1596
               ||||||
Db    531  GlyPheGlyProGlyGlyLeuTyrCysLysAsnCysSerThrTyrSerThrAsnIleAsp  550

Qy   1597  ACATTGCCGCTGGTGGTCAGAGGATGTCACG--------------------CTG  1629
                 :::||||||||  ||||::||||:|||||
Db    551  ArgAlaLysIleLeuTrpGlyTrpAspAlaThrMetThrTyrGlnThrGlnTrpPheAsnLeu  570

Qy   1630  TCACTGAATTACACCTGGACCCGAAGTGAACACACTGATAGGTCATAACAAAGGTGCGCCG  1689
                 ||||||   |||||  :::  :::|||
Db    571  GlyLeuAlaTyrAsnArgThrArgGlyLysAsnGln-----------AsnThrAsnGluTrp  587

Qy   1690  CTGAGTTATACCCCTGAACACATGGTGAATGCGAAACTGAACTCGGCAGATCACCGAAGAG  1749
               |||  :::  :::
Db    588  LeuAspThrIleAsnProAspThrValThrSerThrLeuAspValProValAlaAsnSer  607

Qy   1750  ------GTGGCATCATGGCTGGGGTGCCCGTTATCGCGGGAAAACACCAGTTTCACCCAG  1803
               |||
Db    608  GlyPheAlaValGlyTrpIleGlyThr----------------PheAlaAsp  619

Qy   1804  AATTATTGGTCACTGAGCGGCTGTACAGAGAAGTGTATGATGAGAAAGGAGAATACCTG  1863
               |||  :::||||||:
Db    620  ArgGlySerArgValSerSer-------------------SerGlyThrProGln  631

Qy   1864  AAAGCCTGGACGGTGGTGGATGCGAAGTCTGTGCTGGGAAGATGACGGAT----------GCC  1914
                 :::|||::||||:|  :::
Db    632  LeuAspThrIleAsnProAspThrValThrSerThrLeuAspValProValAlaAsnSerGlyGlyGlnPheLysGlyGly  651

Qy   1915  CTGACGCTGAATGCTGCGTGAATAACCTGCTCAACAAGGATTAC  1959
               :::|||       ::  |||
Db    652  MetThrThrThrValValLeuGlyLysAsnAlaPheAspLysLysGlyTyr  666
```

```
RESULT 9
HMUR_YERPE
ID   HMUR_YERPE     STANDARD;      PRT;    676 AA.
AC   Q56989;
DT   01-NOV-1997 (Rel. 35, Created)
```

DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hemin receptor precursor.
GN   HMUR OR YPO0283 OR YO543.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM6;
RX   MEDLINE=96386041; PubMed=9026634;
RA   Hornung J.M., Jones H.A., Perry R.D.;
RT   "The hmu locus of Yersinia pestis is essential for utilization of
RT   free haemin and haem-protein complexes as iron sources.";
RL   Mol. Microbiol. 20:725-739(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
CC   -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC   UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC   THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U60647; AAC64866.1; -.
DR   EMBL; AJ414141; CAC89146.1; -.
DR   EMBL; AE013655; AAM84131.1; -.
DR   PIR; AG0035; AG0035.
DR   PIR; T12069; T12069.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW   Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW   Complete proteome.
FT   SIGNAL   1   28   POTENTIAL.
FT   CHAIN   29   676   HEMIN RECEPTOR.
FT   SITE   44   51   TONB BOX.
FT   SITE   659   676   TONB C-TERMINAL BOX.
SQ   SEQUENCE   676 AA; 74230 MW; 84ED731CB914ACD3 CRC64;

```
Alignment Scores:
Pred. No.:                    7.95e-17      Length:                  676
```

```
Score:                  341.00      Matches:        169
Percent Similarity:     38.04%      Conservative:   103
Best Local Similarity:  23.64%      Mismatches:     267
Query Match:            9.14%       Indels:         176
DB:                                 Gaps:           31

US-10-625-972-4 (1-2091) x HMUR_YERPE (1-676)

Qy    55  AGCAGCGATAGCTGCTCGCAGCAGATGTGATGATTGTTCGCGCATCCGCTATGAGAAAAAG  114
              |||||    :::    ||        |||:||:|:|||:||::::||||:|||:|
Db    37  SerSerLysHisSerThrAspThrMetValValThrAlaThrArgGlyAsnGluArgSer   56

Qy   115  CTGACTAACGGACGCCAGTGTTCTGTGATTAGCCAGGAGAATTGCCAGTCCAGCCAG     174
              :::||||   ::||| ||||    |||||:|||||||||:||||      |||
Db    57  SerPheGluAlaProMetMetValThrValIleGluGlyAsnAlaProThrSerGlnThr   76

Qy   175  TACCACGATCTGGCCGAGGGCTCTGAGATCAGTAGGAGGGTTGTGAAGTGGTACG       234
              |||||      :::  |||||| |||||::|||||||:||:::|||||||||
Db    77  AlaAlaThrAlaAlaAspMetLeuArgGlnValProGlyLeuThrVal---ThrGlySer  95

Qy   235  GGTAAAACCGAGGGCTGGAAATCAGCATCCGAGAATGCCAGCCAGTTACACCGTGATA    294
             |:|:||| :::::: ::|||||||      |||||| ||||||| ||||||| |||
Db    96  GlyArgThrAsnGlyIleAspValValMetArgGlyTyrGlyLysGlnGlyValLeuThr  115

Qy   295  CTGATTGATGGTGTTCGTCAGGGCGGAAGCAGTGACGTGACTCCCAACGGTTTTCTGCC   354
             |||:|||   :::||||||   ::::  |||
Db   116  LeuValAspGlyValArgGlnGly-----------------ThrAspThrGlyHis----  128

Qy   355  ATGAATACCCGGTTCATGCCCCCTCTGGCCGCCATTGAGCGTATTCAGGTTATCAGGGGG  414
             ::|||:|||::|::|||:  |||       |||||||||||:|:|||||||||||
Db   129  LeuAsnSerThrPheLeuAspPro--AlaLeuValLysArgIleGluIleValArgGly   147

Qy   415  CCGATGTCCACACTGTATGGCTCTGATGGCGGTGTGTTGGTGAATATCATTACCAGA     474
             ||||||||||||||||||||||      |||:||||||||::|||||::|
Db   148  ProAlaAlaLeuLeuTyrGlySerGlyAlaLeuGlyGlyValIleLeuAlaTyrGluThrVal  167

Qy   475  AAGAATGCACAAATGGCTCTCTTCCGTCAATGCAGGGCTGAATCTGCAGGAAAGCAAC    534
             |||||||  :::  :::::  ||||:|
Db   168  AspAlaAlaAspMetLeuGlnProGlyGlnAsnSerGlyTyrArgValTyrArgSerSerAla  187

Qy   535  AAATGGGGTAACAGCAGACCCAGTTAATTCTGGAGCAGTGGT------------CCCCTT  582
             ||:::|||      :::||| |||
Db   188  AlaThrGlyAspHisSer---PheGlyLeuGlyLeuAlaSerAlaPheGlyArgThrAspAsp  206

Qy   583  GTGGATGAGTTCTGTCAGGCTACGCC----GGTAGCACAACAGCAGGGTCAGGGT       636
             :::||||:|||| |||  |||       |||||||:||||||||||||||
Db   207  LeuAspSerGlyIleLeuSerPheGlyThrArgAspLeuAlaAsnIleArgGlnSerAsnGly  226

Qy   637  TCATCGTCACATCGCTGATACAGCAGGGCGATCACGCGGCCTATTCCTATCCCACGAGTCA  696
             :::     ||||||:::     |||||||:|::
Db   227  PheAsnAla------------------ProAsnAspGlu                       233

Qy   697  CAGAATTATAATCTTGGTGCACGTCTTGACTGGAAGGCGTCGGAGCCAGGATGTGCTCTGG  756
             ||||||   ||:::  |||:::| :::|||||||||
Db   234  ThrIleSerAsnValLeuAlaLysGlyThrTrpGlnIleAspSerIleGlnSerLeu---  252

Qy   757  TTTGATATGGATACCACCCGCAGCGTTATAACCGGGATGG-------------------  801
               :::| |||||||||
Db   253  --------SerAlaAsnLeuArgTyrTyrAsnAsnSerAlaIleGluProLysAsn     268

Qy   802  -----------------------------CAACTGGGAGTCTGACGGGGGGATAT  828
                                         :::
Db   269  ProGlnThrSerAlaProSerSerThrAsnValMetThrAsnArgSerThrIleGlnArg  288

Qy   829  GACCGGACCCTGCGCTATGAGCGAAACAAAATTTCAGCTGCTATGATCATACTTTCACC   888
             :::||||||        ||||:::||||||||||:::
Db   289  AspAlaGlnLeuArgTyrAsnIleLysProLeu----AspGlnGluTrpLeu         304

Qy   889  TTCGGAACATGGAAATCGTATCTGAACGGACGAGACAGAGAAAATAAAGGTCGTGAGCTT  948
             :::      |||:::||:::||||||||||:::|||||||||||||||:::
Db   305  AsnAlaThrAlaGlnValTyr------TyrSerGluValGluLeuAsnAlaArg------  320

Qy   949  GTACGCAGTGTACTGAAGCGCGACAAAATGGGGCGTTGCCGGTCAGCAAATGACCGGATGCC 1008
             :::    |||       ||||||||::::||||:::|||||||:::   :::||||||:::
```

(continued — right-hand alignment block)

```
Db   321  -------------------------ProGlnGlySerAlaGluGluGlyArgGluGlnThr  332

Qy  1009  GAATCGAACCTTATCCTG-----AATTCATTACTGCTTACCCCTCTGGGAGAATTCAT  1062
             :::||     ||| |||     :::||||||
Db   333  ThrGluGlyValLysLeuGluAsnArgThrArgLeuPheIleGluSerProAlaSerHis  352

Qy  1063  CTGGTTACGGTGGGGGCGGAGTTCAGAGCTCGTCCATGAAAGACGGAGTTGTCCTTGCC  1122
             |||:||||   ::: |||||||
Db   353  LeuLeuThrTyrGlyThrGlyThrTyrLysGlnGluGlnThrProGly-----------  368

Qy  1123  AGCACAGGTGAAACTTTCCGGCAG------------AAAAGCTGTCGGTATTGCTGACGAT  1173
             |||:::|||||||||||:::||:::
Db   369  GlyAlaThrGluSerPheProGlnAlaLysLysArgPheSerSerGlyTrpLeuGlnAsp  388

Qy  1174  GAGTGGCATCTCACGGAT---GCACTTGCGTGACTGCGGGGCCAGCCGCTATGAAACATCAT  1230
             :::|||||||||||||      :::||||:::|:::
Db   389  GluIleThrLeuArgAspLeuArgProValSerIleLeuLeuAlaGlyThrArgTyrAspAsnTyr  408

Qy  1231  GAGCAATTCGGGAGACACTTCAGTCCGGTGCATATCTGGTCTGGGATGTG--GCAGAT  1287
             :::              |||||||
Db   409  SerGlySerSerAspGlyTyrAla-------------AspValAspAlaAsp  421

Qy  1288  GCCTGGACGCTGAAAGGGCGTGACC--                                1314
             |||:::    :::::|:::
Db   422  LysTrpSerSerArgGlyAlaIleSerIleThrProThrAspTrpLeuMetLeuPheGly  441

Qy  1315  ------------ACGGGATATAAGGCACCCAGAATGGGGCAGCTA---------CAT  1350
                         :::|||||||||||
Db   442  SerTyrAlaGlnAlaPheArgAlaProThrMetGlyGluMetTyrAsnAspSerLysHis  461

Qy  1351  AAAGGGATTAGTGTGTGTCCGGGCAGGAGAAAAACAAATCTA---CTTGGTAACCCCGAC  1407
             :::|||||||||||||||||||||||||||||    :::
Db   462  PheAlaIleProIleArgProGlyLeuThrLeuThrAsnTyrTrpValProAsnProAsn  481

Qy  1408  CTGAAACCCGGAA------GAGAGCGTCAGTTATGAGGCGTGGGGTGTATTACGATAACCCC  1461
             :::||||||||||      |||:::    :::   :::
Db   482  LeuLysProGluThrArgAspGluThrArgIleGlyGluTyrGlyPheGlyLeuArgPheSerLeu  501

Qy  1462  GCCGGTTAACATGCCAATGGCCAGTGCACCAGTTTTATGACTGACTTCTCCAACAAGATTGTCTCT  1521
             |||||||||   :::|||:::
Db   502  LeuMetAlaGluLeuAspAspLeuGlnPheLysValSerTyrPheAsnThrLysAlaLysAsp  521

Qy  1522  TAT--------------------TCCATAAATGATAACACCAATAGCTATGTAAACAGCCGGAAAG  1566
             |||                    :::     |||
Db   522  TyrIleSerThrArgValAspMetGlnAlaMetThrThrMetThrThrSerValAlaAsnIleAspGln  541

Qy  1567  GCCCGGTTGCACGGTGTGGAATTTGCCGGGACACATTGCCGTGTGGTCAGAGGAGTGTCACG  1626
             :::||:::   :::  |||:::||:::
Db   542  AlaLys--------------------------IleTrpProGlyTyrTrpAspAlaSer  550

Qy  1627  CTGTCA--                                                   1659
             :::|||
Db   551  MetSerTyrArgLysThrValPheAsnTrpLeuPheAsnTrpAspLeuAlaTyrAsnArgThrArgGlyLys  570

Qy  1660  CAACGTGATGGTGAT---------------------AACAAAGGTGCGCCGCTGAGTTATACC  1701
             |||||:::|||:::
Db   571  AsnGlnAsnThrAspGluTrpLeuAspThrIleLeuAspThrValThrSerIleVal  590

Qy  1702  CCTGAACACACTGGTGAATGCGAAATGCCAGATCACCGAAGAGGTGGCATCATGG  1761
             :::|||||||:::||||:::
Db   591  AspValProValAlaAlaAsnSerGlyPheSer-----------------ValGlyTrp  603

Qy  1762  CTGGGT---------------------GCCCGGTTATCCGCGGGAAAAACACCAGTTTCACCCAG  1803
             :::|||
Db   604  IleGlyThrPheAlaAsnArgSerSerArgValSerSerSerThrProGln-------  620

Qy  1804  AATTATTCGTCACTGAGCGGTCTGTACAGAGAAGAAAGTGTATGATGAGAAAGAGAATACCTG  1863
             :::
Db   620  ----------------------------------------------------------  620

Qy  1864  AAAGGCCTGGACGGCTGGTGGATGCAGCGGTTCGTCTGCTCGTCGGAAGATGACCGGATGCC  1914
             :::  :::||||:||||:::
```

```
Db   621 AlaGlyTyrGlyValAsnAspPheTyrValSerTyrLysGlyGlnGluAlaPheLysGly 640
Qy  1915 CTGACGCTGAATGCTCGCGGTGAATAACCTGCTCAACAAGGATTAC 1959
              :::||                  ||| |:::||
Db   641 MetThrThrMetLeuLeuGlyAsnValPheGluLysGluTyr 655
```

RESULT 10
FYUA_YERPE
```
ID   FYUA_YERPE     STANDARD;      PRT;   673 AA.
AC   P46359;
DT   01-NOV-1995 (Rel. 32, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Pesticin receptor precursor (IRPC).
GN   FYUA OR PSN OR YPO1906 OR Y2404.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6/69;
RA   Rakin A., Heesemann J.;
RT   "Analysis of the pesticin receptor from Yersinia pestis: role in
RT   iron-deficient growth and possible regulation by its siderophore.";
RL   J. Bacteriol. 177:1824-1833(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM6;
RX   MEDLINE=95204350; PubMed=7896707;
RA   Fetherston J.D., Lillard J.W., Perry R.D.;
```

```
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=6/69;
RA   Buchrieser C., Rusniok C., Couve E., Frangeul L., Billault A.,
RA   Kunst F., Carniel E., Glaser P.;
RT   "DNA sequence of the 102 kbases unstable region of Yersinia pestis.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
```

```
CC   -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
CC       SIDEROPHORE YERSINIABACTIN.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- INDUCTION: BY IRON STARVATION. IRON REGULATION MEDIATED THROUGH
CC       THE FUR PROTEIN.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
```

```
DR   EMBL; Z35104; CAA84487.1; -.
DR   EMBL; U09530; AAA69906.1; -.
DR   EMBL; AL031866; CAA21395.1; -.
DR   EMBL; AJ414150; CAC90722.1; -.
DR   EMBL; AE013845; AAM85962.1; -.
DR   PIR; A56148; A56148.
DR   PIR; AF0232; AF0232.
DR   HSSP; P05825; 1FEP.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec.1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW   Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW   Transport; TonB box; Complete proteome.
FT   SIGNAL     1     22       POTENTIAL.
FT   CHAIN     23    673       PESTICIN RECEPTOR.
FT   SITE      30     37       TONB BOX.
FT   SITE     657    673       TONB C-TERMINAL BOX.
FT   CONFLICT 484    484       G -> D (IN REF. 1).
FT   CONFLICT 514    514       R -> G (IN REF. 1).
SQ   SEQUENCE   673 AA;   73782 MW;  9C39E6010EBCEB2C CRC64;
```

```
Alignment Scores:
Pred. No.:             1.02e-16          Length:         673
Score:                 339.50            Matches:        172
Percent Similarity:    39.14%            Conservative:   120
Best Local Similarity: 23.06%            Mismatches:     283
Query Match:           9.10%             Indels:         171
DB:                                      Gaps:            37
```

```
US-10-625-972-4 (1-2091) x FYUA_YERPE (1-673)

Qy    1 ATGGCGAATAACC-----------------ACTCTGGCTTCCGTAGTCATTCCTGTCTCGGA 45
          ||:::::||||                    ||:|||:|:|| ::::
Db    1 MetLysMetThrArgLeuTyrProLeuAlaLeuGlyGlyLeuLeuLeuProAlaIleAla 20

Qy   46 TTTTCAGCCAGCAGCATAGCTGCTCCAGAGGATGTGATGATTGCTCCGGCATCCGGCTAT 105
          :::        ::|||:|:|||:||||||   :::::|||||
Db   21 --AsnAlaGlnThrSerGlnGlnAspGlnSerThrLeuValThrAlaSerLysSerGln 39

Qy  106 GAGAAAAAGCTGACTAACGCA--GCCGCCAGTGTTCTCGTGATTACGCCAGGAGGAATTG 162
          :::  ::||||                  ::||:|||||||||||
Db   40 SerSerArgSerAlaSerAlaAsnAsnValSerSerThrValValSerAlaProGluLeu 59

Qy  163 CAGTCCAGCCAGTACCACGATCTCGGCGAGGCTCTCGAGATCAGTAGAGGGTGTGGATGTT 222
          ||  |||  :::  ::::|||::::::||||
Db   60 SerAspAlaGlyValThrAlaSerAspLysProArgValLeuProGlyLeuAsnIle 79

Qy  223 GAAAGTGGTACGGGTAAAACCGAGGGCTGAAATCAGCATCCAGGAATGCCAGCCAGT 282
          |||:::||||                  :::::||||||:::::||||::
Db   80 GluAsn---SerGlyAsnMetLeuPheSerThrIleSerLeuArgGlyValSerSerAla 98

Qy  283 TAC------------ACGCTGATACTGATTGATGGTGTTCGTCAGGCGCGGAAGC 324
          :::     :::|||:|||||||
Db   99 GlnAspPheTyrAsnProAlaValThrLeuTyrValAspGlyValProGln----- 115

Qy  325 AGTGACGTGACTCCAACGGTTTTCTGCCATGAATACCGGGTTCATGCCCCTCTGGCC 384
          |||||||||||||||||
Db  116 --------------------LeuSerThrAsnThrIleGlnAlaLeuThr 125

Qy  385 GCCATTGAGCCGATTGAGGTTATCAGGGGGCCGATGTCCACACTGTATGCGCTGATGCG 444
          :::::  :::||:::||||||||||
Db  126 AspValGlnSerValGluLeuArgGlyProGlnInGlyThrLeuTyrGlyLysSerSerAla 145

Qy  445 ATGGGCGGTGGTGGTAATATCATTACCAGGAAAGAATGCAGACAAAATGGCTCTTTCCGTC 504
          |||||||:::|||::|||||||
Db  146 GlnGlyGlyIleIleAsnIleValThrArgGlnInProAspSerThrProArgGlyArgIle 165
```

```
Qy   505 AATGCAGGCTCGAATCTGCAGGAAAGCAACAAATGGGGTAACAGCAGTTAATTTC  564
            |||||||      ::    ||||:||||||
Db   166 GlyGlyValSerSerArgAspSerTyrArg----------SerLysPheAsnLeu  181

Qy   565 TGGAGCAGTGGTCCCCTTGTGGATGATTCTCAGCGTGCGGCGTAGCACACA  624
             ||||||||      :::    ||||||
Db   182 -----SerGlyProIleGlnAspGlyLeuLeuTyrGlySerValThrLeuArgGln  199

Qy   625 CAGGGTCAGGGTTCATCGGTC----------ACATCACTGAGCGAGTACAGGCACGCGT  675
             |||||||||:||:            ||||||:|
Db   200 ValAspAspMetIleAsnProAlaThrGlySerAspSerAspLeuGlyGlyThrArg  219

Qy   676 ----------ATTCCTTATCCCACCGGAGTCACAGAATTATAATCTTGGT  714
                      ::    |||    ||:::||
Db   220 AlaSerIleGlyAsnValLysLeuArgLeuAlaProAspAspTrpGlnMetGly  239

Qy   715 ------GCACGTCTTGACTGGAAGGCGTCGGAGCAGGAT------GTGCTCTGGTTTGAT  762
                   |||     ||||||       |||    |||
Db   240 PheAlaAlaSerArgArgGluCysThrArgAlaThrGlnAspAlaTyrValGlyTrpAsnAsp  259

Qy   763 ATG----------GATACCACCCGGCAGCGTTATGATAACCGGGAT  798
          :::         |||     :::      ||||||
Db   260 IleLysGlyLysLeuSerIleSerArgAspGlySerProAspTyrMetArgArgCys  279

Qy   799 GGGCAACTGGGAGGTGGAGTCGACGGGGGATATGACCGACCCTGGCTATGGAGCGAAACAA  858
          :::||||||:|||           :::    |||
Db   280 ThrAspSerGlnThrLeuSerGlyLysTyr-----ThrThrAspAspValPheAsnLeu  298

Qy   859 ATTTCAGCT----------GGCTATGATCATACTTTCCACCTTCGGAACATGGAAATCG  906
            |||||        ||:::||
Db   299 IleSerAlaTrpGlnGlnHisGlnTyrSerArgThrPheProSerGlySer----------  315

Qy   907 TATCTGAACTGGAACGAGAGACAGAAAATAAAGGTCGTGACCAGTACTACTGAAG  966
                |||||        :::      ||:::|||
Db   316 ----------LeuIleValAsnMetProGln  322

Qy   967 CGGGACAAATGGGGGCTTGCGGTCAGCCGCGGGGAGCTTAAGGAATCGAACCTTATCCTG  1026
          |||      |||         ::    ::||
Db   323 Arg------Trp----------AsnGlnAspVal  328

Qy  1027 AATTCATTACTGCTTACCCCCTCTGGAGAATCTCATCTGGTT---ACGGTGGGGGGCGAG  1083
          |||||||:|:::         |||||:|
Db   329 GlnGluLeuArgAlaAlaThrLeuGlyAspAlaArgThrValAspMetValPheGlyLeuLeu  348

Qy  1084 TTTCAGAGCTCGTCC----------ATGAAAGACGGAGTTGTC  1116
          ::::::  :::::::
Db   349 TyrArgGlnAsnThrArgGluLysLeuAsnSerAlaTyrAspMetProThrMetProTyr  368

Qy  1117 CTTGCCAGCACAGTGAAACTTTCCGGCAGAAAAGTCGGTCGGTATTGCTGGAGGATGAG  1176
          |||||:||||||             :::  ||||||::
Db   369 LeuSerSerThrGlyThrTyrThr---ThrAlaGluThrAlaAlaTyrSerAspLeuThr  387

Qy  1177 TGGCATCTCACGGATGCATGCCGGATGCTGCCGGCAGCCGGCTATGAACATCATGAGCAA  1236
          |||||||||||:||           :::
Db   388 TrpHisLeuThrAspArgPheAspIleGlyLysGlyLysValArgPheSerHisAspLysSer  407

Qy  1237 ----------TTCGGG  1242
                    |||||||
Db   408 SerThrGlnTyrHisGlySerMetLeuGlyAsnProPheGlyLysAspGlnGlyLysSerAsn  427

Qy  1243 ----------GGACACTTCAGTCCGCTGCATATCGGTCTTGGGATGTCGGCAGAT  1287
                    ||||:|||||       :::      |||
Db   428 AspAspGlnValLeuGlyGlnLeuGlnLeuSer---AlaGlyTyrMet-----LeuThrAsp  443

Qy  1288 GCCTGGACGCTGAAAGGCGTGAACCCCAGAATGGGGCAGCTA  1347
          ||||||||||||             ||||||
Db   444 AspTrpArgValTyrArgThrArgValAlaGlnLeuGlyTyrTyrLysPro----------  457

Qy  1348 CATAAAGGGGATTAGTGGTGTGTCCGGGCAGGAGAAAAACAAAATCTACTTGGTAACCCCGAC  1407
          :::  |||             :::
Db   458 --SerGlyTyrAsnIleValProThrArgAlaGlyLeu------------------AspAsp  469

Qy  1408 CTGAAGCCG----------GAAGAGAGCCGTCAGTTATGAGGCGTGGGGTGTATTACGATAAC  1458
```

```
Qy                                                                  489
          |||||
Db   470 AlaLysProPheValAlaAlaGlyThrArgGluLeuAsnTyrGluLysSerIleAsnTyrGluThr  489

Qy  1459 CCCGCCGGTCTGATGCCAAGTTTTATGACTGACTTCTCCAACAAGATTGTC  1518
          :::                |||              |||
Db   490 ---AlaAspValThrLeuGlnAlaAlaThrPheTyrThr---HisThrLysAspMetGln  507

Qy  1519 TCTTATTCCATAAATGATAATCACCAATAGCTATGTAAACAGCGGAAGGCCCGGTTGCAC  1578
          |||||||||||:||
Db   508 LeuTyrSerGlyTyrProValArgMetGlnThrLeuSerAsnAlaGlyLysAlaAspAlaThr  527

Qy  1579 GGTGTGGAATTTGCCGGCACA----------TTGCCGCTGTGTCAGAGGATGTCACGCGT  1629
          |||||||:||:||:            |||||:||
Db   528 GlyValGluLeuGlyLeuAlaLysProAlaProGlyTrpSerTrpAspIle----------  545

Qy  1630 TCACTGAATTACACCTGGACCCGAAGTGAACACGTGATGGTGATAAC------AAAGGT  1683
          :::||||||:::              ||:::|
Db   546 ------AsnGlyAsnValIleArgSerGluPheThrAsnAspSerGluLeuTyrHisGly  563

Qy  1684 GCGCGCGCTGAGTGTATACCCCTGAACACATGGTGAATCGAAACTGACTGGCAGATCACC  1743
          :::            :::          |||
Db   564 AsnArgValProPheValProArgTyrGlyAlaGlySerSerValAlaSerGlyValIle---  582

Qy  1744 GAAGAGGTGGCATCATGGCTGGTGCCCGTATCGGGGGAAACACCACCGGTTCACCCAG  1803
          :::::                  |||
Db   583 ----------AspThrArgTyrGlyAlaLeuMetProArgLeuValAlaVal  595

Qy  1804 AATTATTCGTCACTGAGCGCGTGTACAGAAGAGAAAGTGTATGATGAGAAAGGAGAATCCTG  1863
          ::: |||                ::::|||
Db   596 ------Asn----------LeuValGlyProHisTyrPheAspGlyAspAsnGlnLeuAsnLeuArg  610

Qy  1864 AAA----GCCTGGACGCGTGGTGGATGCAGGTCTGTCGTGGAAGATGACGGATGCCCTGACG  1920
          :::       ::::||||           :::  :::||
Db   611 GlnGlyThrTyrThrAlaThrLeuAspSerSerLeuGlyTrpGlnAlaThrGluArgMetAsn  630

Qy  1921 CTGAATCGTGCGGTGAATAACCTGCTCAACAAGGATTACAGTGACGTGAGCCTGTACAGT  1980
          |||||||:||||              :::      ::::||
Db   631 IleSerValTyrValAspAsnLeuPheAspArgArgTyrArgThrTyrGly--TyrMet  649

Qy  1981 GCCGGTAAGGACGACGCTG  1998
          :::|||  |||
Db   650 AsnGlySerSerAlaVal  655
```

```
RESULT 11
FYUA_YEREN
ID   FYUA_YEREN     STANDARD;      PRT;    673 AA.
AC   P46360;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Pesticin receptor precursor (IRPC) (IPR65).
GN   FYUA.
OS   Yersinia enterocolitica.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=630;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 51871 / WA-314 / Serotype O:8;
RX   MEDLINE=95075311; PubMed=7984105;
RA   Rakin A., Saken E., Harmsen D., Heesemann J.;
RT   "The pesticin receptor of Yersinia enterocolitica: a novel virulence
RT   factor with dual function.";
RL   Mol. Microbiol. 13:253-263(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1223-75-1, 8081 / Serotype O:8, YE737, and 1209-79;
RA   Rakin A., Heesemann J.;
RL   Submitted (JUL-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
CC       SIDEROPHORE YERSINIABACTIN.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
```

```
DR   EMBL; Z29675; CAA82773.1; -.
DR   EMBL; Z35485; CAA84620.1; -.
DR   EMBL; Z35486; CAA84621.1; -.
DR   EMBL; Z35487; CAA84622.1; -.
DR   EMBL; Z35496; CAA84630.1; -.
DR   PIR; S60142; S60142.
DR   HSSP; P05825; 1FEP.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW   Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW   Transport; TonB box.
FT   SIGNAL     1     22       POTENTIAL.
FT   CHAIN     23    673       PESTICIN RECEPTOR.
FT   SITE      30     37       TONB BOX.
FT   SITE     657    673       TONB C-TERMINAL BOX.
FT   VARIANT  362    362       N -> D (IN STRAINS 8081, YE737 AND 1209-
FT                             79).
SQ   SEQUENCE   673 AA;  73682 MW;  1C20E0352FAD4DCC CRC64;

Alignment Scores:
Pred. No.:              3.26e-16          Length:        673
Score:                  332.50            Matches:       172
Percent Similarity:     39.08%            Conservative:  118
Best Local Similarity:  23.18%            Mismatches:    289
Query Match:            8.91%             Indels:        163
DB:                     1                 Gaps:          36

US-10-625-972-4 (1-2091) x FYUA_YEREN (1-673)

Qy    1 ATGGGAATAACC-----------ACTCTGGCTTCCTGCTAGTCATTCCCTGTCTCGGA  45
           ::   :::::                :::::::::: :::
Db    1 MetLysMetThrArgLeuTyrProLeuAlaLeuGlyLeuLeuLeuProAlaIleAla    20

Qy   46 TTTTCAGCCAGCAGCATAGCTGCTGCAGAGGATGTGATGATTGTCTCGGCATCCGGCTAT 105
           :::  :  ::: :::    |||||
Db   21 --AsnAlaGlnThrSerGlnGlnAspGluSerThrLeuValThrAlaSerLysLysGln   39

Qy  106 GAGAAAAAGTCGACTAACGCA---GCCGCCAGTGTTCTGTGATTAGCCAGGAGGAATTG  162
           :::                    :::::::
Db   40 SerSerArgSerArgAlaSerAlaAsnValSerSerThrValValSerAlaProGluGlu  59

Qy  163 CAGTCCAGCCAGTACCACGATCTGGCGGAGGAGGCTCGACATCAGTCAGTGGTGGGATGTT 222
           :::         ||||    :::::::
Db   60 SerAspArgAlaGlyValAlaThrAlaSerAspLysProArgValLeuProGlyLeuAsnIle 79

Qy  223 GAAAGTGGTACGGGTAAAACCGGAGGGCTGGAAATCAGCATCCCAGGAATGCCAGCCAGT  282
          |||::   :::::            :::::::
Db   80 GluAsn--SerGlyAsnMetLeuPheSerThrLeuArgGlyValSerSerAla          98

Qy  283 TAC----------ACGCTGATACTGATTGATGGTGTTCGTCAGGGCGGAGC--------  324
          :::          :::::::::::   :::::::
Db   99 GlnAspPheTyrAsnProAlaValThrLeuTyrValAspGlyValProGln--------- 115

Qy  325 AGTGACGTGACTCCCAAGCGTTTTTCTGCCATGAATAACGGGTTCATGCCCCTGGCC    384
                    :::        |||
Db  116 ------------LeuSerThrAsnThrIleGlnAlaLeuThr                   125

Qy  385 GCCATTGAGCGTATTGAGTTATCAGGGGGCCGATGTCCACACTGTATGGCTCTGATGCG   444
          ::::     :::||||:::::|||
Db  126 AspValGlnSerValGluLeuArgGlyProGlnGlyTyrLeuTyrGlyLysSerAla    145

Qy  445 ATGGGCCGGTGTGTCCGGCAGGGAAAAACAAATCACTTGGTAACCCGACCTGAAGCCG   504
          :::
Db  146 GlnGlyGlyIleIleAsnIleValThrGlnAsnProAspSerThrProArgGlyTyrIle 165

Qy  505 AATGCAGGGCTGAATCTGCAGGAAGCAACAAATGGGTAACACAGCAGCCAGTTAATTTC  564
          :::::   ::::::           ::   |||:||||
Db  166 GluGlyGlyValSerSerArgAspSerTyrArg--------SerLysPheAsnLeu     181

Qy  565 TGGAGCAGTGGTCCCCTTGTGGATGATTCTGCAGCTGGAGGTACGCGGTAGCACACAA   624
          ::::::||||
Db  182 -----SerGlyProIleGlnAspGlyLeuLeuTyrGlySerValThrLeuValArgGln  199

Qy  625 CAGCGTCCAGGGTTCATCGGTC---------ACATCACTGAGCGATACAGCAGCACGCGT 675
          |||                   ::
Db  200 ValAspAspGlyAspMetIleAsnProAlaThrGlyLysAspAspLeuGlyGlyThrArg  219

Qy  676 -------------------ATTCCTTATCCCACGGAGTCACAGAATTATAATCTTGGT   714
                             :::   :::||||
Db  220 AlaSerIleGlyAsnValLysLeuArgLeuAlaProAspAspGlnTrpGluMetGly    239

Qy  715 -------GCACGTCCTGACTGAAGGCGTCCGGACCAGGAT------GTGCTCGGTTTGAT 762
                 ||||        :::      :::
Db  240 PheAlaAlaSerArgGluCysThrArgAlaThrGlnAlaTyrValGlyTrpAsnAsp    259

Qy  763 ATG-----------------------GATACCACCCGGCAGCGTTATGATAACCGGGAT  798
          :::                        |||    :::
Db  260 IleLeuLysGlyArgLysLeuSerLeuSerAspGlySerProAspTyrMetArgArgCys 279

Qy  799 GGGCAACTGGGAGTCTGACGCGGGGGATATGACCGGACCCTGCGCTATGAGCGAAACAAA 858
          :::                                     |||
Db  280 ThrThrAspSerGlnThrLeuSerGlyLysTyr---ThrThrAspAspPheAsnLeu    298

Qy  859 ATTTCAGCT-----------GGCTATGATCATACTTCACCTCCGAACATGGAAAATCG   906
          :::||||||          |||:::        ||||||
Db  299 IleSerAlaTrpGlnGlnGlnHisTyrSerArgThrPheProSerGlySer------Leu 316

Qy  907 TATCTGAACTGGAACGAGCAGCAGAAAATAAAGGTCGTGAGCTTGTACGCAGTGTACTGAAG 966
          :::||||   :::    ::  ::::::
Db  317 IleValAsnMetProGlnAspTrpAsnGlnAsnValGlnLeuLeuArgAlaAlaThrLeu   336

Qy  967 CGGCAGACAAA------------------TGGGGCTTGCCGGTCAGCCGGGGAGCTTAAG 1008
          :::                          ::::::||||
Db  337 GlyAspAlaArgThrValAspMetValPheGlyLeuTyrArgGlnAsnThrArgGluLys  356

Qy 1009 GAATGCAACCTTATCCTGAATTCATTACTGCTTACCCCTCTCGGGAGAATCCATCTGTT  1068
          ||||||||||
Db  357 ------------------------LeuAsnSer                            359

Qy 1069 ACGGTGGGGGGGCCAGTTCAGAGCTCGTCCATGAAAGACGGAGTTGTCCTTGCCAGCACA 1128
          ::||||||          :::|||
Db  360 ----------------------AlaTyrAsnMetProThrMetProTyrLeuSerSerThr 372

Qy 1129 GGTGAAACTTTCCGGCAGAAAAGCTGGTGGTATTGCTGAGGATGGTGGCATCTCACG   1188
          |||                         |||
Db  373 GlyTyrThr---ThrAlaGluThrLeuAlaAlaAlaTyrSerAspLeuThrTrpHisLeuThr 391

Qy 1189 GATGCACTTGCGCTGACTGCGGGCAGCCGCTATGAACATCATGAGCA------------ 1236
          |||:::                                  :::
Db  392 AspArgPheAspIleGlyGlyGlyValArgPheSerHisAspLysSerThrGlnTyr   411

Qy 1237 ------------------------TTCGGG------                         1242
                                  TTCGGG
Db  412 HisGlySerMetLeuGlyAsnProPheGlyLeuGlyLysSerAsnAspAspGlnVal   431

Qy 1243 --GGACACTTCAGTCCGTCCCATATCGGGTCTCGGGGATGTGCAGATGTGGACGCTG   1299
            ||||||||                           :::
Db  432 LeuThrAspAspTrpArgVal                                        447

Qy 1300 AAAGGCGGTGTGACCACGGGGATATAAGGCCGTACATAAAGGGATT              1359
          ||||||||
Db  448 TyrThrArgIleAlaGlnIleGlyTyrLysPro--------SerGlyTyr          460

Qy 1360 AGTGGTGTGTCCGGCAGGGAAAAACAAATCACTTGGTAACCCGACCTGAAGCCG---    1416
          :::                                                 |||
```

```
Db    461 AsnIleValProThrAlaGlyLeu-------------AspAlaLysProPhe 473

Qy   1417 -----GAAGAGAGGCTCAGTTATGAGGCTGGGGTGTATTACGATAACCCGCCGGTCTG 1470
                ||:::|||| |||:::||||||| ||||  |||  :::
Db    474 ValAlaGluLysSerIleAsnTyrGluLeuGlyThrArgTyrGluThr---AlaAspVal 492

Qy   1471 AATGCCAATGCACAGGTTTATGACTCTCCAACAAGATTGTCTTATTCCATA 1530
               ||| ||  ||| :::  :::     ||||||
Db    493 ThrLeuGlnAlaAlaThrPheTyrThr---HisThrLysAspMetGlnLeuTyrSerGly 511

Qy   1531 AATGATAACACCAATAGTCATGTAAACAGCGGAAAGGCCCGGTTGCACGGTGGGAATTT 1590
               :::              ||||||||||||||||
Db    512 ProValGlyMetGlnThrLeuSerAsnAlaGlyLysAlaAspAlaThrGlyValGluLeu 531

Qy   1591 GCCGGCACA-------TTGCCGCTGTGTGTCAGGATGTCACGCTGTCACTGAATTAC 1641
               |||||          |||:::|| |||:::|
Db    532 GluAlaLysTrpArgPheAlaProGlyTrpSerTrpAspIle---------AsnGly 547

Qy   1642 ACCTGACCCGAAGTGAACACAGTCGATGGTGATAAC------AAAGGTGCGCCCCTGAGT 1695
               ||||||||||          :::  :::
Db    548 AsnValIleArgSerGluPheThrAsnAspSerGluLeuTyrHisGlyAsnArgValPro 567

Qy   1696 TATACCCCTGAACACATGGTGAATGCGAAACTGAACTGCCAGATCACCCGAAGAGGTGGCA 1755
               :::  |||     :::   ||  ||| |||:::  ||
Db    568 PheValProArgTyrGlyAlaGlySerSerValAsnGlyValIle---------582

Qy   1756 TCATGGCTGGGTGCCCGTTATCGCGGGAAAACACCACGTTCACCCAGAATATTCGTCA 1815
               |||||||          ||  ||||||
Db    583 --------AspThrArgTyrGlyAlaLeuMetProArgLeuAlaValAsn---------596

Qy   1816 CTGAGCGCCTGTACGAAGAAAGTGTATGAGACAAGAGGAATACCTGAAA---GCCTGG 1872
               :::||||          ||:||| |||
Db    597 -----LeuValGlyProHisTyrPheAspGlyAspAsnGlnLeuArgGlnGlyThrTyr 614

Qy   1873 ACGGTGGTGGATGCGAGCTGTCGTGGAGATGACGGATGCCCCTGACGCTGAATGCTGCG 1932
               ||:|||| |||   |||:::||| |||  |||:::
Db    615 AlaThrLeuAspSerSerLeuGlyTrpGlnAlaThrGluArgIleAsnIleSerValHis 634

Qy   1933 GTGAATAACCTGCTCAACAAGGATTACAGTGACGTGAGCCTGTACAGTGCCGGTAAGAGT 1992
               |||||          |||:::|||  |||
Db    635 ValAspAsnLeuPheAspArgArgTyrArgThrTyrGly---TyrMetAsnGlySerSer 653

Qy   1993 ACGGTG 1998
               :::|||
Db    654 AlaVal 655
```

```
RESULT 12
YBIL_ECOLI
ID   YBIL_ECOLI     STANDARD;     PRT;     760 AA.
AC   P75780;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable tonB-dependent receptor ybiL precursor.
GN   YBIL OR B0805.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
```

```
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [3]
RP   IDENTIFICATION BY MASS SPECTROMETRY, AND PUTATIVE FUNCTION.
RX   MEDLINE=20267815; PubMed=10806384;
RA   Molloy M.P., Herbert B.R., Slade M.B., Rabilloud T., Nouwens A.S.,
RA   Williams K.L., Gooley A.A.;
RT   "Proteomic analysis of the Escherichia coli outer membrane.";
RL   Eur. J. Biochem. 267:2871-2881(2000).
CC   -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT, THAT PARTICIPATES IN
CC       IRON TRANSPORT.
CC   -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AE000182; AAC73892.1; -.
DR   EMBL; D90717; BAA35471.1; -.
DR   EMBL; D90718; BAA35476.1; -.
DR   PIR; E64817; E64817.
DR   EcoGene; EG13317; ybiL.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep.Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
KW   Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW   Complete proteome.
FT   SIGNAL        1     33       POTENTIAL.
FT   CHAIN        34    760       PROBABLE TONB-DEPENDENT RECEPTOR YBIL.
FT   SITE        743    760       TONB C-TERMINAL BOX.
SQ   SEQUENCE   760 AA;  81960 MW;  1B66B9C2EF68C7CA CRC64;
```

```
Alignment Scores:
Pred. No.:                              5.82e-14    Length:              760
Score:                                  301.50      Matches:             172
Percent Similarity:                     36.36%      Conservative:        120
Best Local Similarity:                  21.42%      Mismatches:          258
Query Match:                            8.08%       Indels:              253
DB:                                     1           Gaps:                40

US-10-625-972-4 (1-2091) x YBIL_ECOLI (1-760)

Qy     37 TGTCTCGGATTTCA---------GCCAGCAGCATAGCGCT----------GCAGAG 75
             ||||:||| |        :::          ||| ||||||         |||:::
Db     22 CysIleGlyIleThrProValAlaGlnAlaLeuAlaLeuAlaGluGlnThrAsnAlaAsp 41

Qy     76 GATGTGATGATTGTCTCGGCATCC---------------------------- 99
             ||||||||| |||||||||||||
Db     42 AspThrLeuValValGluAlaSerThrProSerLeuTyrAlaProGlnSerAlaAsp 61

Qy    100 ---GGCTATGAGAAAAAGCTGACTAACGCCACCCGCCAGTGTTCTGATTAGCCAGGAG 156
               :::|||||| || |||                  ||  ||
Db     62 ProLysPheSerArgProValAlaAspThrThrArgThrMetThrArgThrGluGln 81

Qy    157 GAATTGCAGTCCAGCCAGTACCACCGATCTGCGGAGGCTCTGAGATCAGTAGAGGGTGTG 216
             :::::: |||   |||||         :::||| :::|||
Db     82 ValIleLysAspGlnGlyAlaThrAspLeuThrAsnLeuLysAsnValProGlyVal 101

Qy    217 GATGTT-----------GAAAGTGGTACGGGTAAAACCCGAGGGCTGAAATCAGCATC 264
             :::||||         :::|||          ||||||
```

```
Db  102  GlyAlaPhePheAlaGlyGluAsnSerThrThrGlyAsp--AlaIleTyrMet 120
Qy  265  CGAGGAATGCCAGCCAGTTACACGCTGATACTGATTGATGGTGTTCGTCAGGGCGGAAGC 324
              ||||||   :::    |||    |||||||||:::|||    ||||||
Db  121  ArgGlyAlaAlaAspThrSerAsnSer---IleTyrIleAspGlyIleArgAspIleGlySer 139

Qy  325  AGTGACGTGACTCCCAACCGTTTTTCTGCCATGAAATACCGGGTTCATGCCCCCTCTGGCC 384
              ::::::||||| |||     |||||   |||||| ::||
Db  140  ------ValSerArgAspThrPhe----------AsnThr-------------- 147

Qy  385  GCCATTGAGCGTATTGAGGTTATCAGGGGGCCGATGTCCACACTGTATGGCTCTGATGCG 444
              ||:::::|||||||||||   |::||||   ||||||
Db  148  ------GluGlnValGluValIleLysGlyProSerGlyThrAspTyrGlyThrArgSerAla 165

Qy  445  ATGGGCGGCGTGGTGGTGAATATCATTACCAGAAAG--------AATGCA 483
              ||||||||   :::::::  ::::::||
Db  166  ProThrGlySerIleAsnMetIleSerGlnProArgAsnAspSerGlyIleIleAspAla 185

Qy  484  GACAAATGGCTCTCTTCCGTC-- 504
              |||   |||   |||
Db  186  SerAlaSerIleGlySerAlaTrpPheArgArgGlyThrLeuAspValAspAspGlnValIle 205

Qy  505  --------------------------------------AATGCAGGGCTG 516
                                                   :::|||||||
Db  206  GlyAspThrThrAlaValArgLeuAsnValMetGlyGluLysThrHisAspAlaGlyArg 225

Qy  517  AATCTGCAGGAAAGCAAATGGGGT--------AACAGCAGCCAGTTTAATTCTGGAGC 570
              ::   :::::   ||||||||          |||   |||   |||   |||
Db  226  AspLysValLysAsnGluIleArgTyrGlyValAlaProSerValAlaProPheGlyLeuTyrThr 245

Qy  571  AGTGGGTCCCCTGTG------------------------GATGATTCT 594
              :::   |||
Db  246  AlaAlaAsnArgLeuTyrLeuAsnTyrLeuHisValThrGlnHisAsnThrProAspProGlyGly 265

Qy  595  GTC--AGCCTGCAGGTACGCGGTAGCCACCACACAGCGTCAGGGTCATCGGTCACATCA 651
              :::   |||   :::    ||    |||
Db  266  IleProThrIleGlyLeuProGlyTyrSerGlyGlyAlaProSerAlaGlyIleThrAlaAlaAlaLeuAsn 285

Qy  652  CTGAGCGATACAGCACAGCCACGCGTATTCCTTATCCACCGGAGTCACAGAATTATAATCTT 711
              |||   |||   :::   :::|||   ||    |||:::||
Db  286  HisSerGlyLysValAspThrHisAsnPheTyrGlyThrAspSerAsp-------------- 301

Qy  712  GGTGCACGTCTTGACTGGAAGGCGTCGGAAGGCAGGATGTGCTCCTGGTTTGATATGGAT-- 768
              |||       :::   ||    :::   :::||
Db  302  ----------TyrAspAspSerThrHisAlaThrMetArgPheGluHisAspIle 318

Qy  769  -------------ACACCCGGCCAGCGTTATGATAACCGGGATGGGCAA 804
                           ||||||||||
Db  319  AsnAspAsnThrIleArgAsnThrThrArgTrpSerArgValLysGlnAspTyrLeu 338

Qy  805  CTGGGGGAGTCGACGGGGGATATGACCCGACCCTGCGCTATGAGCGGAAACAAAATTCA 864
              :::   :::::   ||||||
Db  339  MetThrAlaIleMetGlyThrValAlaSerAsnIleThrGln-------------- 351

Qy  865  GCTGGCTATGATCATCACTTTCACCTTCGGAAATCGGAAATCGTATCTGAACACGAGCAG 924
              ||:::::::::   |||:::
Db  352  ----------ProThrSerAspValAsnSerTrp-------ThrTrpSerArg 363

Qy  925  ACAGAAAATAAAGGTCGTGAGCCTTGTACGCAGTGTACTGAAGCCGGACAAAATGGGGGCTT 984
              |||   :::::::   |||   :::
Db  364  ThrAlaAsnThr---LysAspValSerAsnLysIleLeu------------ 375

Qy  985  GCCGGTCAGCCCGCGGGAGCTTAAGGAATCGAACCTTATCCTGAATTCATTACTGCTTACC 1044
              :::   :::::::   |||   :::::::
Db  376  --------ThrAlaSerGlnInThrAsnLeu---ThrSerThrPheTyrThrGly 388

Qy  1045 CCTCTGGGAGAATCCATCCGGTTACGGTACCGGAGTT---------- 1086
              :::||     |||:::   |||||||
Db  389  SerIleGly-----HisAspValSerThrGlyValGluPheThrArgGluThrGlnThr 406

Qy  1087 --------------------------------------CAGAGCTCG 1095
                                                   |||||||
Db  407  AsnTyrGlyValAsnProValThrLeuProAlaValAsnIleTyrHisProAlaSerSer 426
```

```
Qy  1096 TCCATGAAAGACGGAGTTGTCCTTGCCGACCACAGGTGAAACTTTCCGGCAGAAAAGCTCG 1155
              |||   :::
Db  427  IleHisProGlyGlyLeuThrArgArgAsnGlyAlaAsnAlaAsnGlyGlnThrAspThrPhe 446

Qy  1156 TCGGTATTGCTGAGGATGAGTGGCATCTCACGGATGCACTTGCCTGACTGCCGGGCAGC 1215
              :::::::|||   |||    ::::||
Db  447  AlaIleTyrAlaPheAspThrLeuGlnIleThrArgAspPheGluLeuAsnGlyGlyIle 466

Qy  1216 CGCTATGAA--CATCATGAGCAATTC-------GGGGGACACTTCAGT 1254
              |||   :::||:::   |||
Db  467  ArgLeuAspAsnTyrHisThrGluTyrAspSerAlaThrAlaCysGlyGly-----Ser 484

Qy  1255 CCGGGTGCATATCTGTCTCGGATGTGGCAGATGCCTGCAGACGCGTGAAAGGCGGTGTGACC 1314
              ||||||           |||||||         ||||||
Db  485  GlyArgGlyAlaIleThrCysProThrGlyValAla--------LysGlySerProVal 501

Qy  1315 ACGGGATATAAGGCACCCAGAATGGGCAGCTACATAAAGGATTAGTGGT------- 1365
              :::                              |||:::   :::|||
Db  502  ThrThrValAspThrAlaLysSerGlyAsnLeuMetAsnTrpLysAlaGlyAlaAlaLeuTyr 521

Qy  1366 ------------------------------------------GTGTCCGGGCAG 1377
                                                       |||||||
Db  522  HisLeuThrGluAsnGlyAsnValTyrIleAsnTyrAlaValSerGlnGlnProProGly 541

Qy  1378 ----------------GGAAAAACAAATCTACTTGGTAACCCCGACCTG 1410
                               |||       :::
Db  542  GlyAsrAsnPheAlaLeuAlaAlaGlnSerGlyAsnSerAlaAsnArgAsnArgAspPhe 561

Qy  1411 AAGCCCGAAGAGAGCGTCAGTTATGAGCGTGGGGTGTATTAC--------GATAACCCC 1461
              |||||||       :::   :::||   |||   |||
Db  562  LysProGlnLysAlaAsnThrSerGlnIleGlyThrLysTrpGlnValLeuAspLysArg 581

Qy  1462 GCCGGTCCTGAATGCCAATGCCAAGTTTATGACTGACTTCTCCAACACAGATTGTCTCT 1521
              ||||||       |||:::        |||||||:::::::
Db  582  LeuLeuLeuThrAlaAlaLeu------PheArgThrAspIleGluAsnGluValGluGln 599

Qy  1522 TATTCCATAAATGATAACACCAATAGCATGTAAACACGCGAAAGGCCCGGTTGCACGGT 1581
              |||||||||
Db  600  --------AsnAspAsp--GlyThrTyrSerGlnInTyrGlyLysLysArgValGluGly 615

Qy  1582 GTGGAATTT-----GCCGGCACATTG---CCGCTGGTGGTCAGGAGGATGTCACGCTGTCA 1632
              |||||||||        |||   :::      ||:::   :::Val  Ile 631
Db  616  TyrGluIleSerValAlaGlyAsnIleThrProAlaTrpGln---------ValIle 631

Qy  1633 CTGAATTACACCTGGACCCGAAGTGAACAACGTGATGGT-----------GATAACAAA 1680
              :::   :::::   |||||||
Db  632  GlyGlyTyrThrGlnInLysAlaThrIleLysAsnGlyLysAspValAlaGlnGlyIle 651

Qy  1681 GGTGCCGCCTGAGTTATACCCCGAACACATGTGAAATGGAAACTGAACTGGCAGATC 1740
              |||||||||||||||||       :::::::
Db  652  SerSerSerLeuProTyrThrProGluHisAlaPheThrLeuThrSerGlnTyrGlnAla 671

Qy  1741 ACCGAAGAGGTGGCATCATCGCTGCGGTGCCCGTTATCGCGGAAA-------- 1785
              |||::::::::::   ||||||||||||
Db  672  ThrAspAspIleSerValGlyAlaGlyAlaArgTyrIleGlySerMetHisLysGlySer 691

Qy  1786 ----------------ACACCAGGTTCACCCAGAATATTCGTCACTGAGCGCGTGTACAG 1830
                               ||||||:::::::   :::
Db  692  AspGlyAlaValGlyThrProAlaPheThrGluThrGlyTyr------------- 704

Qy  1831 AAGAAAGTGTATGATGAAGAAGGAGAATACCTGAAAGCCTGGACGTGTGGATGCAGGT 1890
              :::                              |||   :::   |||
Db  705  -----------------Trp--ValAlaAspAlaLys 710

Qy  1891 CTGTCGTGGAAGATGACGGATGCCCTGACGCGTGAATAGTGCGTGAATAACCTGCTCAAC 1950
              :::::::::::          :::::::
Db  711  LeuGlyTyrArgValAsnArgAsnLeuAspPheGlnLeuAsnLeuTyrAsnLeuPheAsp 730

Qy  1951 AAGGATTAC 1959
              |||||||||
Db  731  ThrAspTyr 733
```

```
RESULT 13
FECA_ECOLI
ID   FECA_ECOLI     STANDARD;      PRT;    774 AA.
AC   P13036;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Iron(III) dicitrate transport protein fecA precursor.
GN   FECA OR B4291.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 34-40.
RC   STRAIN=B;
RX   MEDLINE=88227855; PubMed=2836368;
RA   Pressler U., Staudenmaier H., Zimmermann L., Braun V.;
RT   "Genetics of the iron dicitrate transport system of Escherichia
RT   coli.";
RL   J. Bacteriol. 170:2716-2724(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=95334362; PubMed=7610040;
RA   van Hove B., Staudenmaier H., Braun V.;
RT   "Novel two-component transmembrane transcription control: regulation
RT   of iron dicitrate transport in Escherichia coli K-12.";
RL   J. Bacteriol. 172:6749-6758(1990).
RN   [3]
RP   SEQUENCE OF 1-428 FROM N.A.
RX   MEDLINE=91072220; PubMed=2254251;
RA   van Hove B., Staudenmaier H., Braun V.;
RT   "Nucleotide sequences of the fecBCDE genes and locations of the
RT   proteins suggest a periplasmic-binding-protein-dependent transport
RT   mechanism for iron(III) dicitrate in Escherichia coli.";
RL   J. Bacteriol. 171:2626-2633(1989).
RN   [4]
RP   SEQUENCE OF 731-774 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=89213950; PubMed=2651410;
RA   Staudenmaier H., van Hove B., Yaraghi Z., Braun V.;
RT   "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT   region from 92.8 through 100 minutes.";
RL   Nucleic Acids Res. 23:2105-2119(1995).
RA   Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA   Blattner F.R.;
CC   -!- FUNCTION: FECA IS THE OUTER MEMBRANE RECEPTOR PROTEIN IN THE
CC       IRON(III) DICITRATE TRANSPORT SYSTEM.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- INDUCTION: FOR INDUCTION THE TONB AND THE EXBB PROTEIN HAVE TO
CC       ACTIVE. REGULATION BY THE IRON LEVEL MEDIATED BY THE FUR PROTEIN
CC       AND INDUCTION BY CITRATE PLUS IRON SUGGEST THAT THE IRON(III)
CC       DICITRATE COMPLEX MUST ENTER THE PERIPLASM WHERE IT BINDS TO A
CC       TRANSMEMBRANE PROTEIN, WHICH REGULATES DIRECTLY OR VIA A FURTHER
CC       INDUCTOR, TRANSCRIPTION OF THE FEC GENES.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; M20981; AAA23760.1; -.
DR   EMBL; U14003; AAA97187.1; -.
DR   EMBL; AE000499; AAC77247.1; -.
DR   EMBL; M63115; AAA23768.1; -.
DR   EMBL; M26397; AAA23761.1; -.
DR   PIR; E65242; QRECFA.
DR   PDB; 1KMO; 23-OCT-02.
DR   PDB; 1KMP; 23-OCT-02.
DR   EcoGene; EG10286; fecA.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW   Complete proteome; 3D-structure.
FT   SIGNAL        1     33
FT   CHAIN        34    774       IRON(III) DICITRATE TRANSPORT PROTEIN
FT                                FECA.
FT   SITE         56     63       TONB BOX.
FT   SITE        757    774       TONB C-TERMINAL BOX.
FT   CONFLICT     16     16       T -> A (IN REF. 1).
FT   CONFLICT    190    190       A -> T (IN REF. 1).
FT   CONFLICT    300    300       T -> M (IN REF. 1 AND 3).
FT   CONFLICT    357    357       W -> R (IN REF. 1).
FT   CONFLICT    444    444       L -> V (IN REF. 1).
FT   CONFLICT    749    749       R -> A (IN REF. 1).
SQ   SEQUENCE   774 AA;  85321 MW;  16B5B510276C3B09 CRC64;

Alignment Scores:
Pred. No.:                              774
Score:                  6.91e-14        Length:             774
Percent Similarity:     300.50          Matches:            169
Best Local Similarity:  37.52%          Conservative:       109
Query Match:            22.81%          Mismatches:         282
DB:                     8.05%           Indels:             181
                                        Gaps:               40

US-10-625-972-4 (1-2091) x FECA_ECOLI (1-774)

Qy      64 GCTGCTGCAGGAGGATGTGATGATTGTC----------TCGGCATCCGGCTATGAGAAA  111
              ||||||||   ::|       |||||                |||||
Db     109 AlaProLysGluAspGlyAlaLeuThrValValGlyAspTrpLeuGlyAspAlaArgGluAsn  128

Qy     112 AAGCTGACTAACGACCAGCCGCCCAGTGTTTCTGATTAGCCAGGAGGAATTGCAGTCCAGC  171
              :::    |||||| |||  :::  |||||   ||||  ||| |||:||
Db     129 AspValPheGluHisAlaGlyAlaArgAspValIleArgArgGluArgGlyAlaLysThr  148

Qy     172 CAGTACCACGATCTCGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTT------GAA  225
              :::|||  ||| |||  |  |||| :::   |||:|
Db     149 GlyAlaThrThrMetArgGluValLeuAsnArgIleProGlyValSerAlaProGluAsn  168

Qy     226 AGTGGTACGGTAAAACCGGA---GGCTGGAAATCAGCATCCGAGGAATGCCA--------  276
              :::||||||          |||::::
Db     169 AsnGlyThrGlySerHisAspLeuAlaMetAsnPheGlyIleArgGlyLeuAsnProArg  188

Qy     277 ---GCCAGTTACACGCTGATACTGATTGATGGTGTTCGTCAGGGCGGAAGCAGTGACGTG  333
                 |||| |  | :::||:::|||||::
Db     189 LeuAlaSerArgSerThrValLeuMetAspGlyIle--------------ProValProPhe  204

Qy     334 ACTCCCAACGTTTTTCTGCCATGAATACCGGGTTCATGCCCCT-----CTGGCGCGC  387
              |||      ||  ||||  |||||||||||||  :::  ||
Db     205 AlaProTyrGlyGlnProGlnLeuSer------------LeuAlaProValSerLeuGlyAsn  221

Qy     388 ATTGAGCGATTGAGGTTATCAGGGGGGCCGATGCCACACTGATGGCTCGATGCGATG  447
              :::::  |||||::|     ||||| ||||||| :::  :::::
Db     222 MetAspAlaIleAspValAlaArgGlyGlyValAlaValArgTyrProGlnSerVal  241

Qy     448 GGCGGTGTGGTGAATATCATTACCAGAAAAGAATGCAGACAAATGGCTCTTCCGTCAAT  507
              ||||||||||||||    ||||||||||      ||   :::
Db     242 GlyGlyValValAsnPheValThrArgAlaIleProGlnAspPhe------GlyIleGlu  259

Qy     508 GCAGGGCTGAATCTGCAGGAAAGCAACAAATGGGTAACAGCAGCCAGTTAATTTCTGG  567
              ||:::  |||    ||||||||| ||     |||||||:::||||
Db     260 AlaGlyValGluGlyGlnLeuSer-----------ProThrSerSerGlnAsnAsn------  274

Qy     568 AGCAGTGGTCCCCTTGTCGTGATGATTCTGTCAGCCTGCAGGTACGCGGTAGCACCAACAG  627
              |||||||||||||| :::||||||||  |||||||||
Db     275 ------------ProLysGluThrHisAsnLeuMetValGlyThrAlaAspAsn  289

Qy     628 CGTCAGGGTCATCGGTACACTGAGCGATACAGCAGCCAGTTATTCCTTATCCC  687
              |||  :::|| ||||||     :::
Db     290 GlyPheGlyThrAlaLeuLeu--TyrSerGlyThrArgGlySer---------  303
```

```
Qy  688  ACGGAGTCACAGAATATAATCTTGGTGCACGTCTTGACTGGAAG-------GCGTCG    738
              ||||||||                              |||::|||
Db  304  ------------------------------------------------AspTrpArgGluHisSerAlaThr    311

Qy  739  GAGCAGGATGTGCTCTGGTTT----------------------------GATATGGAT    768
              ||| |||                                      :::
Db  312  ArgIleAspAspLeuMetLeuLysSerLysTyrAlaProAspGluValHisThrPheAsn    331

Qy  769  ACCACCCGGCAGCGTTATGATAACCGGATGGGCAACTGGGAGTCTG-----------    816
              :::  ::::::|||:::|||
Db  332  SerLeuLeuGlnTyrTyrAspGlyGluAlaAspMetProGlyGlyLeuSerArgAlaAsp    351

Qy  917  ----------------------ACGGGGGGATATGACCGGACCCTGCGCTATGAGCGA    852
                                    |||||||||
Db  352  TyrAspAlaAspArgTrpGlnSerThrArgProTyrAspArgPhe-----TrpGlyArg    369

Qy  853  AACAAAATT--TCAGCTGGCTAT--------------------GATCATACTTCACCTTC    891
              ||||::    |||  ||||||                    |||  |||
Db  370  ArgLysLeuAlaSerLeuGlyTyrGlnPheGlnIleProAspSerGlnHisLysPheAsnIle    389

Qy  892  --GGAACATGGAAATCGTATCTGAACCGGACGAGACAGAAAATAAAGGTCGTGAGCTT    948
                ||| :::                |||::||||| ::
Db  390  GlnGlyPheTyrThrGlnThrLeuArgSerGlyTyrTyrLeuGlu---GlnGlyLysArgIle    408

Qy  949  GTACGGCAGTGTACTGAACGGCGACAAATGGGGGCTTGCCGTCAGCCGCGGGAGCTTAAG    1008
              |||                                          ::::||||||
Db  409  ThrLeuSer-------ProArgAsnTyrTrpValArgGlyIleGluProArg--------    423

Qy  1009  GAATCGAACCTTATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAATCT---CATCTG    1065
              |||                                  :::|||||||   |||
Db  424  TyrSerGlnIlePheMet-----------------------IleGlyProSerAlaHisGlu    436

Qy  1066  GTTACGGTGGGGGGCCAGAGTTTCAGAACTCGTCCATGAAAGACGGAGTTGTCCTTGCCAGC    1125
              |||  |||||||      :::  ::: |||  |||
Db  437  ValGlyValGlyTyrArgTyrLeuAsnGluSerThrHisGluSerArgThrTyrThrAla    456

Qy  1126  ACAGGTGAAACTTTCCGGCAGAAAAGCTGGTCGGTATTTGCTGAGGAT----------    1173
              :::                                             |||
Db  457  ThrSerSerGlyGlnLeuProSerGlySerSerProTyrAspArgArgAspThrArgSerGly    476

Qy  1174  ----------GAGTGGCAATCCACGGATGCACTT---------------GGCGTGACT    1206
                        |||||:::||| :::               :::|||
Db  477  ThrGluAlaHisAlaTrpTyrTyrLeuAspAspLysIleAspGlyAsnTrpIleThr    496

Qy  1207  GCGGGGCAGCCGCTATGAACATCATGAAGCAATTC-----------------GGGGGACAC    1248
              ||| |||::::||||||| :::                       ||||||
Db  497  ProGlyMetArgPheGluHisIleGluSerTyrGlnAsnAsnAlaIleThrHisGlyThrHis    516

Qy  1249  TTCAGT----------------------------CCGCGTGCATATCTGGTCTCGGGATGTGGCAGAT    1287
                                                |||        ::: |||
Db  517  GluGluValSerTyrAsnAlaProLeuProAlaLeuAsnValLeuTyrHisLeuThrHisAsp    536

Qy  1288  GCCTGGACGCGTGAAAGGCGGTGTGACCAGGCAGGGCACCCAGAATGGGGCAGCTA    1347
              ::||||  |||                :::                |||||:::
Db  537  SerTrpAsnLeuTyrAlaAsnThrGluGlyTyrSerPheGlyTyrValGlnTyrSerGlnIle    556

Qy  1348  CATAAAGGGATTAGTGCGTGTCCCGGCACAGGAGAAAACAAATCTACTTGGTAACCCCGAC    1407
              ||||||||||      :::::||||||
Db  557  GlyLysLysAlaValGlnSer----------------------GlyAsn--------    564

Qy  1408  CTGAAGCCGGAAGAGAGCCGTCAGTTATGAGGCGTGTATTACGATAACCCCGCCGGT    1467
              :::::::||||::::: :::::::||| |||  |||||||:::
Db  565  ValGluProGluLysAlaArgThrArgGlnLeuLeuGlyIleGlnThrArgTyrAspAspGlyAla--    583

Qy  1468  CTGAATGCCAATGTCACAGGTTTTATGACTGACTTCTCCAACAAGATTGTCTCTTATCC    1527
              |||                |||                  :::
Db  584  LeuThrAlaGluMetGlyLeuMetGlyLeuPheLeuIleLeuAsnPheAsnAsnGln-----------TyrAsp    600

Qy  1528  ATAAATGATAACACCATTAGCTATGTAAACGCCGGAAAGGCCCCGGTTGCACGGGTGTGAA    1587
              |||                                |||::::||
Db  601  SerAsnGlnThrAsnAspThrValThrAlaArgGlyLysThrArgHisThrArgGlyLeuGlu    620
```

```
Qy  1588  TTT-------------------GCCGGCACATTGCCGCTGGTCTGTGGTCAGGAGGATGTCACGCTGTCA    1632
                                   ||||||||        :::::||
Db  621  ThrGlnAlaArgTyrAspLeuThrLeuTthrProThrLeuAspAsnValSerIleTyr    640

Qy  1633  CTGAATTACACCTGACCCGAAGTGAACACACGTGAT---GGTGATAACAAAGGTGCGCCG    1689
              :::|||    :: ||||                   ||||||
Db  641  AlaSerTyrAlaTyrAlaAsnAlaGluIleArgGluLysGlyAspThrTyrGlyAsnLeu    660

Qy  1690  CTGAGTTATACCCCTGAACACATGGTGAATGCGAAACTGAACTGGCAGATCACCGAAGAG    1749
              :::  ::::|||:::||
Db  661  ValProPheSerProLysHis----------------------------    667

Qy  1750  GTGGCATCATGCTGGTGCCCGTTATCGCGGGGAAAACACCACGTTTCACCCAGAATTAT    1809
                               |||               :::
Db  668  ---LysGlyThrLeuGlyValGlyAspThrTyrLysProGlyAsnTrpThrPheAsnLeuAsn---    685

Qy  1810  TCGTCACTGAGCGCTGTACAGAAGAAAGTGTATGATGAGAAAGGAGAAATACCTGAAAGCC    1869
              |||                               :: :::       ::||
Db  686  -------SerAspPheGlnSerSerGlnPheAlaAlaAspAsnThrValLysGlu    702

Qy  1870  TGGACGGTGGTGGATGCAGGT-----------------CGTGCGTGG---------    1899
              ||||||||||||:::|||
Db  703  SerAlaAspGlySerThrGlyArgIleProGlyPheMetLeuTrpGlyAlaArgValAla    722

Qy  1900  ------------------AAGATGACGGAGCGCCTGACGCTGAATCGTGCGGTGAATAACCTG    1944
                                :::|||       ||| ||||
Db  723  TyrAspPheGlyProGlnMetAlaAsp---LeuAsnLeuAlaPheGlyValLysAsnIle    741

Qy  1945  CTCAACAAGGATTACAGTGACGTGAGCCTGTACAGTGCCGGTAAGAGTACGCGTGTATGCC    2004
              :::::::||||||                   :: |||||
Db  742  PheAspGlnAspTyrPhe---IleArgSerTyrAspAspAsnLysGlyLyIleTyrArgAla    760

Qy  2005  GGT    2007
          |||
Db  761  Gly    761
```

RESULT 14

```
Y262_HAEIN
ID   Y262_HAEIN    STANDARD;    PRT;    723 AA.
AC   P44600;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable tonB-dependent receptor HI0262 precursor.
GN   HI0262.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   IDENTIFICATION BY MASS SPECTROMETRY.
RX   MEDLINE=20137488; PubMed=10675023;
RA   Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA   Gray C., Fountoulakis M.;
RT   "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL   Electrophoresis 21:411-429(2000).
CC   -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
```

```
CC   -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; U32712; AAC21927.1; -.
DR   PIR; C64058; C64058.
DR   TIGR; HI0262; -.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_dep_Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Outer membrane; Receptor; Signal; TonB box; Complete proteome.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    723       PROBABLE TONB-DEPENDENT RECEPTOR HI0262.
FT   SITE        706    723       TONB C-TERMINAL BOX.
SQ   SEQUENCE   723 AA;  80775 MW;  91EB3AB0FFEA2984 CRC64;

Alignment Scores:
Pred. No.:                          Length:        723
Score:              3.01e-13        Matches:       166
Percent Similarity: 291.50          Conservative:  116
Best Local Similarity: 36.15%       Mismatches:    275
Query Match:        21.28%          Indels:        223
DB:                 7.81%           Gaps:           32
                    1
```

```
US-10-625-972-4 (1-2091) x Y262_HAEIN (1-723)

Qy       46  TTTTCAGCCAGCAGCATAGCTGCTCGAGAGGATGGATGCTTGTCTCGGCATCCGGCTAT  105
                |||||| |||||:||:|||       :||:|||
Db        3  PheSerLysSerLeuHisLeuThrThrThrLeuValThrAlaAsnAlaLeuAlaGln   22

Qy      106  GAGAAAAAGCTGACTAAGCGAGCCAGCCGCCAGTGTTCTGTAGTT----------------  147
                :::|:||:|||||||||              :||SerIleAsnValIleAlaThrArgAspProSer   38
Db       23  SerValGluLeuAsp---------------

Qy      148  ---------------AGCCAGGAGGAATTGCAGTCCAGTCCAGCCAGTACCAC  180
                                |||:||:|||  ||:|||
Db       39  ArgPheAlaTyrThrProGluLysGlnSerLysAspSerLeuLeuSerLysGlnAlaThr  58

Qy      181  GATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTGAAAGTGTACGGGTAAA  240
                  ::  |||:|||:||:|||              ||:|||::|   :::
Db       59  SerValAlaAspAlaLeuGluAspIleProAsnValAspValArgGlySer---Arg  77

Qy      241  ACCGGAGAGGCTCGGAAATCAGCATCGGAGGAATGCCAGTTACACGGTGATACTGATT  300
                :::      ::|:|||||             :::|| :::|:||
Db       78  SerIleAlaGlnLysProAsnIleArgGlyLeuSerAspAsnArgValValGlnValIle  97

Qy      301  GATGGTGTTCGTCAGGGCGAAGCAGTGACAGTGACCTCCACGGTTTTTCTGCCATGAAT  360
                ||||||:||              ::: ||:|||
Db       98  --------AspPheAspLeuAlaHisArgGly---   110

Qy      361  ACCGGGTTCATGCCCCCTCTGCCGCCATTGAGCGTATTGAGGTTATTCAGGGGGGCCGATG  420
                ::      :::|:|||||             |||||:|||:::
Db      111  --SerTyrPheLeuProMetSerLeuIleGluGlnIleLysGluValIleLysGlyProSer  129

Qy      421  TCCACACTGTATGGCTCTGATGCGATGGCGCGGTGTGTGAATATCATTACC-----------  471
                |||::|||:|||||:||||||||||
Db      130  SerSerLeuTrpGlySerGlyAlaLeuGlyValValAlaMetArgThrProAsnAla  149

Qy      472  ---------------AGAAAGAATGCAGACAAATGGCTCTCTTCCGTCAATGCAGGGGCTG----  516
                                ||||||:|| ||||:||
Db      150  LeuAspLeuLeuLysAsnAsnAspLysPheGlyValLysIleArgGlnLysTyrGlnThr  169

Qy      517  ----------------AATCTCCAGGAA--  528
                                ||||||:||
Db      170  AlaAsnAsnLeuSerGluLysGluLysAspValSerValPheAlaAlaAsnAspLysPheAspVal  189
```

```
Qy      529  ----------------------------------AGCAACAAATGG  540
                                                   |||
Db      190  LeuIleSerGlyPheTyrAsnAsnAlaAspAsnLeuArgGlyLysLysGlyAsnLysLeu  209

Qy      541  GGTAACAGCAGCCAGTTAATTTCTGGAGCAGCAGTGTCCCCTGTGGATGATTCTGTCAGC  600
                |||::|::::|:                 |||   |||
Db      210  AsnAsnThrAlaTyrLysGlnPhe------GlyGlyLeuAla----LysPheGly  224

Qy      601  CTGCAGGTACGCGGTAGCACAACAGCCTCAGGGTTCATCGGTCACATCACTGAGGCGAT  660
                |||  :::  :::   |||
Db      225  TrpGlnIleAsnAspAlaAsnArgValGluLeuSerHisArgGluThrArgPheLysGln  244

Qy      661  ACAGCAGGCACGGTATTCCTATCCCACGGAGTCACAGAATTATAATCTTGTGCACGT  720
                |||||||            :::    :::|||
Db      245  ThrAlaProSerAsnAsnGluValGluAsnGluLeuThrAsnGluGlnIleThrAspGln  264

Qy      721  CTT--GACTGGAAGGGCTCGGAGCAGCAGATGTGTCTCTGGTTTGATATGGATACCACCCGG  777
                :::         ::: |||:::||               :::
Db      265  IleLysLysPheHisGlyGlnLysAspAspLeuLeu-------ProProThrGlnThr  281

Qy      778  -----------------CAGCGGTTATGATAACCGGGATGGGCAACTGGGAGTCTG  816
                                 :::|||||||||:::||||||||||
Db      282  ProSerProSerGluPheTyrSerLysValLysThrArgLeuGlySerVal  301

Qy      817  ACGGGGGGATATGACCGACCCTGCGCTATGAGCAGAACAAAAATTTCAGCTGGCTATGAT  876
                |||:||||              ||||||:::|||
Db      302  SerTyrLeuThrAspGlnGlnIlePro--------------------------AspThr  311

Qy      877  CATACTTTCACCTTCGGAACATGG-----------AAATCCTATCGAACTGG  918
                ||||||:||||||||||||||       ||||||||||||
Db      312  GlnSerThrValPheAsnTyrTyrLeuThrProAspAsnProTyrLeuAsnThrHisIle  331

Qy      919  -----------------AACGAGACAGAAAATAAAGGTCGAGCTTGTACGCAGTGTACTGAAG  966
                                 |||:||::||:|||
Db      332  AlaLeuTyrAsnAsnLysThrIleGluGluLysGluGlnLysArgLysValSerGlyVal---  349

Qy      967  CCGCACAAATGGGGGCTTGCCGGTCAGCCGCCGGAGCTTAAGGAATGCGAACCTATCCTG  1026
                :::|:|::                |||:|||||:::
Db      350  LysAspGlnInThrLysLeuThrLeuThrArgGlyIleAsnLeuArgAsnSerSerGluLeu--  368

Qy     1027  AATTCATTACTGCTTACCCCCTCGGGAGAATCTCATCTGGTTACGGTGGGGGGGC-----  1080
                                                :::|||
Db      369  --------------------SerHisIleSerPheValTyrGlyValValAsp  378

Qy     1080  ----------------------------------------------------------  1080
                
Db      379  TyrMetArgAspLysIleArgThrGluArgGlyThrAsnGlySerAspAlaLysPheArg  398

Qy     1081  -------GAGTTCAGAGCTGCTCCATGAACGGAGTTGTCCTTGCCAGCACAGGTGAA  1134
                       :::   :::|||:||              ||||||
Db      399  AlaAspProTyrAsnAlaAsnSerAsnThrThrGlyValTyrLeuIleAlaHisIlePro  418

Qy     1135  ACTTTCCGGCAGAAAGCTGGTCGGATGATTGCTGCGGATGAGTGCATCCACCGATGCA  1194
                |||                |:::
Db      419  LeuPheGlyGluLys-----------------------------------------  423

Qy     1195  CTTGCGCTGACTCGCGGCAGCCGCTATGAACATGAGCAATTCGGG------------  1242
                |||      :::::                |||:::::
Db      424  LeuLeuValSerProSerValArgTyrAspHisTyrAspThrSerLysThrValLys  443

Qy     1243  -------GGACACTTCAGTCCGTGCATATCGGTCTCGGGATGTGGCAGATGCCTG  1293
                       ||||:||              |||:||
Db      444  TyrLysAspAsnHisLeuSerProAlaThrLysLeuThrTrpIleValThrAsnTrpLeu  463

Qy     1294  ACGTGAAAGCGGTGACCACGGATATAAGGCACCCAGAATGGGGCAG----  1344
                |||:::              |||:||
Db      464  AspPheThrAlaLysTyrAsnGluAlaPheAlaProSerMetGlnArgPheVal  483

Qy     1345  --------CTACATAAAGGATTAGTAGGTGTGTCCCGGCAGGAAAAACAAATCACTTCGT  1398
                        |||   :::  :::
Db      484  SerGlyAlaHisPheGlyAlaAlaAsnThrLeuGlyLeuAspHisIleAsnArgPheValAla  503
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).

DR   EMBL; U09840; AAA87059.1; -.
DR   PIR; A57148; A57148.
DR   InterPro; IPR000531; TonB boxC.
DR   Pfam; PF00593; TonB dep Rec; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR   PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW   Outer membrane; Transport; TonB box; Signal; Receptor.
FT   SIGNAL        1     19       OR 21 (POTENTIAL).
FT   CHAIN        20    725       HEME/HEMOPEXIN UTILIZATION PROTEIN C.
SQ   SEQUENCE   725 AA;  80838 MW;  FC7886E020CB5BCE CRC64;
```

Alignment Scores:
```
Pred. No.:              3.28e-13        Length:          725
Score:                  291.00          Matches:         173
Percent Similarity:     36.27%          Conservative:    115
Best Local Similarity:  21.79%          Mismatches:      259
Query Match:            7.80%           Indels:          247
DB:1                                    Gaps:             33
```

US-10-625-972-4 (1-2091) x HXC2_HAEIN (1-725)

```
Qy   46 TTTTCAGCCAGCAGCATAGCTGCTGCAGAGGATGGTGATGATGATTGTCTCGGCATCCGGCTAT 105
              :::::          ::|::||
Db    3 PheSerLysLeuSerLeuAlaIleAlaThrThrLeuValThrAlaAlaAsnAlaLeuAlaGln 22

Qy  106 GAGAAAAAGTGACTAACGCAGCCAGCCGCCAGTGTTTCTGTGATT------ 147
         |||||||||
Db   23      SerIleAsnValIleAlaThrAlaArgAspProSer 38

Qy  148 ------AGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC 180
               ::::::::::::
Db   39 ArgPheAlaTyrThrProGluLysGlnSerLysSerLeuLeuSerLeuLeuSerGlnAlaThr 58

Qy  181 GATCTGGCGGGAGCCTCTGAGATCGTCGAGAAGTGTGAGGGTGGTGATGTTGAAAGTGGTACGGGTAAA 240
         ||||||||
Db   59 SerValAlaAlaAlaAlaGluAspIleProAsnValAspIleArgGlyGlySer---Arg 77

Qy  241 ACCGGAGGGCTGGAAATCAGCATCGGAGAATGCCAGGAGGTTACACGCCAGTGATACTGATT 300
                 ||||||||
Db   78 SerIleAlaGlnLysProAsnIleArgGlyLeuSerArgSerAsnArgValGlnValIle 97

Qy  301 GATGGTGTTCGTCAGGGCGGAAGCAGCAGTGACGTGACTCCCAACCGTTTTCTGCATGAAT 360
         |||||||||||
Db   98 AspGlyValArgGln------AsnPheAspLeuAlaHisArgGly--- 110

Qy  361 ACCGGGTTCATGCCCCTGCGGCCCCATTGAGCGTATTGAGGGTATCAGGGGGCCGATG 420
         |||||||       :::::::||||||
Db  111 ---SerTyrPheLeuProMetSerLeuHisGlnLysGluGluIleGluValIleLysGlyProSer 129

Qy  421 TCCACACTGTATGGCTCTGATGCGATGGGCGGTGGGTGGTGAATATCATTACC---- 471
          :::|||||
Db  130 SerSerLeuTrpGlySerGlyValAlaLeuGlyValValAlaMetArgThrProAsnAla 149

Qy  472 ------AGAAAGAATGCAGACAAAATGGCTCTTCCTCCGTCAATGCAGGGGCTG--- 516
               |||||||   ||||||
Db  150 LeuAspLeuLeuLysLysAsnAsnAspLysPheGlyValLysIleArgGlnGlyTyrGlnThr 169

Qy  517 ------AATCTGCAGGAA 528
               |||||||
Db  170 AlaAsnAsnLeuSerGluArgAspValSerValPheAlaAlaAsnAspLysPheAspVal 189

Qy  529 ------AGCAACAAATGG 540
Db  190 LeuIleSerGlyPheTyrAsnAsnAlaAsnLeuArgThrGlyLysGlyAsnLysLeu 209

Qy  541 GGTAACAGCAGCCAGTTAATTTCGGAGCAGTGTCCCCTTG---TGGATGATTCTGTC 597
```

```
Qy 1399 AACCCCGGTTGAAGCCGGAGAGAGACGTCAGTTATGAG------GCTGGGTGTATTAC 1452
         |||||||:|||||||||||||||  ::  ||  ::::::       :::::::||
Db  504 AsnProAsnLeuArgProGluThrAlaAlaLysAsnLysGluIleThrAlaAsnLeuHisPhe 523

Qy 1453 GATAAC------CCCGGCCGGTCTGAATGCCAATGCCACAGTTTTATGACTGAC 1500
         ||:||          ||||||||
Db  524 AspSerLeuPheLysGlnGlyAspLysPheLysIleGluAlaThrTyrPheArgAsnAsp 543

Qy 1501 TTCTCCAACAAGATT------GTCTCCTATTCC 1527
         |||                        |||
Db  544 ValLysAspPheIleAsnLeuLysIlePheAsnAspAlaLysThrSerAlaSerAlaGly 563

Qy 1528 ATAAATGATAACACCAAT------AGCTATGTAAACAGCGGA 1563
         ||| |||||||                    |||
Db  564 AlaAsnProAsnThrAsnGlyAlaAlaLeuProLysAsnSerGlnTyrGlnAsnIleThr 583

Qy 1564 AAGGCCCGGTTGCACGGGTGTGGAATTGCCGGCACATTGCCGCTGTCGTCAGGAGATGTC 1623
         ||||||||  |||||||         :::          :::
Db  584 AsnAlaArgLeuSerGlyIleGluLeuGlnAlaGlnTyrGln------ThrGluArgLeu 601

Qy 1624 ACGCTGTCACTGAATTACACCTGGACCCGAAGTGAACAACGTGATGGTGATAACAAAGGT 1683
         ||| ||||||||||      ::  |||||            ||
Db  602 ThrLeuPheThrAsnTyrGlySerThrLysGlyLysAspLysAsp--------SerGly 618

Qy 1684 GCGCCGCTGAGTTATACCCCTGAACACATGGTGAATGCGAAACTGAACTGGCAGATCACC 1743
         |||||||   ::   ::        :: ::   :
Db  619 GluAlaLeuSerAsnIleAlaAlaSerIleGlyIleValGlyIleValAlaAsnTyrAlaLeuVal 638

Qy 1744 GAAGAGGTGGCATCATGGCGTGGGTGCCCTTATCGCGGGAAAACACCACGTTCACCCAG 1803
         ::|:::                     :::|||||
Db  639 LysAsp------------------------------LysPheThrVal 644

Qy 1804 AATTATTCG---TCACTGAGCGCGTGTACAAGAGAAAGTGTATGATGAGAAAGGAGAATAC 1860
         |||||        ::                  |||:::
Db  645 GlyAlaThrValThrHisTyrAlaAlaGlnArgValProLysAspHisSerValThr 664

Qy 1861 CTGAAAGGCTGGACGTGGTGGATGCAGTCTGTCG---------------TGG 1899
         :::::||||||       |||
Db  665 TyrProSerTyrIleLeuThrAspLeuArgAlaThrTyrAlaProLeuLysGlyGluTrp 684

Qy 1900 AAGATGACGGATGCCCTGACCGCTGAATGCTGCGGTGAATAACCTGCTCAACAAGGATTAC 1959
         |||       ||||||||||||
Db  685 Lys------------AsnLeuArgLeuAspPheAlaLeuGluAsnLeuPheAspArgLysTyrTyr 701
```

```
RESULT 15
HXC2_HAEIN
ID   HXC2_HAEIN     STANDARD;      PRT;   725 AA.
AC   P45557;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Heme/hemopexin utilization protein C precursor.
GN   HXUC.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DL42 / Serotype B;
RX   MEDLINE=95270579; PubMed=7751272;
RA   Cope L.D., Yogev R., Mueller-Eberhard U., Hansen E.J.;
RT   "A gene cluster involved in the utilization of both free heme and
RT   heme:hemopexin by Haemophilus influenzae type b.";
RL   J. Bacteriol. 177:2644-2653(1995).
CC   -!- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
CC       CONCENTRATIONS.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC   -!- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE D OF
CC       H.INFLUENZAE.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

Thu Oct 14 08:22:06 2004

us-10-625-972-4.rsp

Page 26

```
Db 514 ThrAlaAsnLeuHisPheAspSerLeuPheLysGlnGlyAspLysPheIleGluAla 533
                                                               |||
Qy 1483 ACAGGTTTATGACTGACTTCTCCAACAAGATTGTCTTATTCCATAAATGAT------ 1536
        |||   |||                         ::: |||
Db 534 ThrThrTyrPheArgAsnAspValLysAspPheIleAsnLeuIlePheAsnAspAlaLys 553

Qy 1537 --AACACCAATAGC------------------------------------------- 1548
          ||||||||||:::
Db 554 ThrAsnThrAsnAlaSerAlaSerAlaLeuAlaGlyAlaAlaGlyAlaAsnProAsnGlyAla 573

Qy 1549 ------------------TATGTAAACAGCGGAAAGGCCCGGTTGCACGGTGTCGGAA 1587
                          |||                   |||  :::|||
Db 574 LeuLeuProThrLysSerGlnTyrGlnAsnIleThrAsnAlaArgLeuSerGlyIleGlu 593

Qy 1588 TTTGCCGGCACATTGCCGGTCGTCGGTCAGGATGTCACGCTGTCACTGAATTACACCTGG 1647
        :::|||          ||||||||||
Db 594 LeuGlnAlaGlnTyrGln------ThrGluArgLeuPheThrAsnTyrGlySer 611

Qy 1648 ACCCGAAGTGAACAACGTGATGGTGATAACAAAGGTGCGCCGCTGAGTTATACCCCTGAA 1707
        ||||::: :::  :::|||                |||
Db 612 ThrLysGlyValGlyValAspLysAsp------SerGlyUlaLeuSerAsnIleAlaAla 628

Qy 1708 CACATGGTGAATGCGAAACTGACTGGCAGATCACCGAAGAGGTGGCATCATGGCTGGGT 1767
        :::::|||    :::::
Db 629 SerLysIleGlyValGlyValAsnTyrAlaLeuValLysAsp------ 642

Qy 1768 GCCCGTTATCGCGGGAAAAACACCACGTTCACCCAGAATTATTCG---TCACTGAGCGCT 1824
                                         ::: |||||
Db 643 --------LysPheThrValGlyAlaThrValThrHisTyrAla 654

Qy 1825 GTACAGAAGAAAGTGTGATGATGAGAAAGGAGAATACCTGAAAGCCTGGACGGTGGTGGAT 1884
        :::||||||          |||
Db 655 AlaGlnArgArgValProLysAspHisSerValThrTyrProSerValThrLeuThrAsp 674

Qy 1885 GCAGGTCGTCG--------------------TGGAAGATGACGGATGCCCTGACGCTG 1923
        ||||||                        |||||
Db 675 LeuArgAlaThrTyrAlaProLeuLysGlyGlyGluTrpLys--------AsnLeuArgLeu 691

Qy 1924 AATGCTGCGGTGAATAAACCTGCTCAACAGGATTAC 1959
        :::  |||::: ||||| |||||  ::::
Db 692 AspPheAlaLeuGluAsnLeuPheAspArgLysTyr 703

Db 210 AsnAsnThrAlaTyrLysGlnPhe-GlyGlyLeuAlaLysPheGlyTrpGlnIleAsnAs 229
       |||  :::::
Qy 598 AGCCTGCAGGTACGCGGTAGCACAC------------ 622
       |||         :::::::|||
Db 229 pAlaAsnArgValGluLeuSerHisArgGluThrArgPheLysGlnThrAlaProSerAs 249

Qy 623 -AACAGCGCGTCAGGGTTCATCGGTCACCATCATCAGCAGGCACGCGTATTCCT 681
        ||||||       :::::::
Db 249 nAsnGluValGlyLeuAsnGluLeuThrAsnGluGlnIleThrAspGlnIleArgGluPhe- 268

Qy 682 TATCCCACGGAGTCACAGAATTATATCTTGGTGACGTCTTGACTGGAAGGCGTCGGAG 741
       |||::  |||             |||      |||::|
Db 269 ------HisLys--ProAsnAsnGlySerProProLysAlaLysProSerGln 283

Qy 742 CAGGATGTGCTCTGGTTTGATATGGATACCACCCGGCAGCGTTATGATAACCGGATGGG 801
       ::::::::
Db 284 GluGluPhe------TyrSerGlyValLysThr 292

Qy 802 CAACTGGGGAGTCTGACGGGGGATGACCGGACCCTGGCGTATGAGCGAAACAAAATT 861
       |||||||::::::          |||::
Db 293 ArgPheGlySerValSerPheTyrLeuThrArgAspGlnIlePro------ 306

Qy 862 TCAGCTGGCTATGATCATACTTTCACCTTCGGAACATGG------------AAATCGTAT 909
       ||| ::: |||  :::                  :::
Db 307 ------------AspGlnSerThrValPheAsnTyrTyrLeuThrProAspAsnProTyr 322

Qy 910 CTGAACTGG------------AACGAGACAGAAAATAAAGGTCGTGAGCTTGTA 951
       ||||||
Db 323 LeuAsnThrHisIleAlaLeuTyrAsnAsnLysThrIleGluLysGlnArgLysVal 342

Qy 952 CGCAGTGTACTGAAGCGCGACAAAATGGGGCGCTTCCGCGTCAGCGCCGGGAGCTTAAGGAA 1011
       |||::::|||          :::
Db 343 SerGlyVal-----LysAspGlnThrLysThrArgGlyIleLeuAsnLeuArgAsn 360

Qy 1012 TCGAACCTTATCCTGAATTCATTACTGCTTACCCCTCGGGAGAATCTCATCTGGTTACG 1071
        ||||||:::
Db 361 SerSerGluLeu----------SerHisIleSerPhe 369

Qy 1072 GTGGGGGGC----------------------------------------GGGGGGC-- 1080
        |||
Db 370 ValTyrGlyValAspTyrMetArgAspLysIleArgThrGluArgGlyThrAsnAsnLys 389

Qy 1081 -----------GAGTTCAGAGCGTCGTCCATGAAAGACGGAGTTGTCCTT 1119
                   ::: :::::|||         ||||||
Db 390 AspAlaGlnPheArgAlaAspProTyrAsnAlaAsnSerAsnThrThrGlyValThrTyrLeu 409

Qy 1120 GCCAGCACAGGTGAAACTTTCCGGCAGAAAAGCTGGTCGGTATTGGTGAGGATGAGTGG 1179
        |||               ::
Db 410 IleAlaHisIleProLeuPheGlyLeuLys-------------- 419

Qy 1180 ------------------CTACATAAAGGATTAGTGGTGTCCGGGCAGGGGAAAA 1239
                          ::: |||         |||::
Db 420 ------LeuLeuLeuSerProSerValArgTyrAspHisTyrAspThrSer 434

Qy 1240 GGG------------------GGACACTTCAGTCCGGTCCATATCTGGTCTCGGGAT 1278
        |||                  ::  |||      |||
Db 435 SerLysThrValLysTyrLysAspAsnHisLeuSerProAlaThrLysLeuThrTrpIle 454

Qy 1279 GTGGGCAGATGCCTGGCGGTGTGAAAGGCCGGTGACCACGGATATAGGCACCCAGAATG 1338
        |||              :::
Db 455 ValThrAsnTrpLeuAspPheThrAlaLysTyrAsnGluAlaPheArgAlaProSerMet 474

Qy 1339 GGGCAG------------CTACATAAAGGATTAGTGGTGTCCGGGCAGGGGAAAA 1383
        :::
Db 475 GlnGluArgPheValSerGlySerHisGlyThrSer---IleLeuGlyValArgAsnGlu 493

Qy 1384 ACAAAT--CTACTTGTAACCCCGACCTGAAAGCCGGAAGAGAGCGTCAGTTATGAG---- 1437
        |||         :::
Db 494 IleAsnLysPheValAlaAlaAsnProAsnLeuArgProGluThrAlaLysAsnLysGluIle 513

Qy 1438 ---GCTGGGGTGTATTACGATAAC------------CCCGCCGGTCTGAATGCCAATGTC 1482
        |||                     :::::::|||
```

Search completed: October 13, 2004, 11:49:51
Job time : 66.5 secs